NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

COMPUTER MODELING OF VOICE SIGNALS
WITH
ADJUSTABLE PITCH AND FORMANT FREQUENCIES

by

Geoffrey T. Hall

December 1978

Thesis Advisor: S.R. Parker

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Computer Modeling of Voice Signals with Adjustable Pitch and Formant Frequencies		5. TYPE OF REPORT & PERIOD COVERED Master's Thesis; December 1978
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Geoffrey Thomas Hall		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Postgraduate School Monterey, California 93940		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Postgraduate School Monterey, California 93940		12. REPORT DATE December 1978
		13. NUMBER OF PAGES 128
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) Naval Postgraduate School Monterey, California 93940		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Linear Predictive Coding
Digital Speech Processing
Frequency Adjustment

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Digital encoding of speech to allow more efficient transmission at low data rates involves the decomposition of the speech waveform into various parameters which are related to the physical structure of the speech production process. In this thesis, linear predictive coding is used to produce a set of coefficients for the characteristic

DD FORM 1473, 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE
(Page 1) S/N 0102-014-6601 |

 SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

polynomial of sucessive 25 msec. segments of the voice tract, in the z-domain. The location of the poles in the z-plane and the excitation pitch period are then shifted and the signal reformulated to cause changes of the overall frequency characteristics of the speech waveform, while maintaining the perceived sounds and information content. The resulting audio tapes confirm the theory and conjectures of the thesis.

Approved for public release; distribution unlimited.

Computer Modeling of Voice Signals with Adjustable Pitch and Formant Frequencies

by

Geoffrey T. Hall
Captain, United States Marine Corps
B.S., Purdue University, 1971

Submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN ELECTRICAL ENGINEERING

from the

NAVAL POSTGRADUATE SCHOOL

December 1978

ABSTRACT

Digital encoding of speech to allow more efficient transmission at low data rates involves the decomposition of the speech waveform into various parameters which are related to the physical structure of the speech production process. In this thesis, linear predictive coding is used to produce a set of coefficients for the characteristic polynomial of sucessive 25 msec. segments of the voice track, in the z-domain. The location of the poles in the z-plane and the excitation pitch period are then shifted and the signal reformulated to cause changes of the overall frequency characteristics of the speech waveform, while maintaining the perceived sounds and information content. The resulting audio tapes confirm the theory and conjectures of the thesis.

TABLE OF CONTENTS

I. INTRODUCTION

Digital processing of speech signals has become important and necessary with the introduction of high-speed digital devices into every phase of communication: place to place; man to machine; and machine to man.

Digital signals have a number of inherent advantages over analog signals. Digital signals may be coded for security or for noise immunity. A digital voice signal may be transmitted by the same equipment used for data and it may be multiplexed with that data. One of the primary disadvantages of the digital transmission of voice is the large bandwidth required with some digital techniques. When analog techniques, such as single side-band amplitude modulation, produce bandwidths of 5KHz and the best digital system bandwidth was 64khz, there was a very strong tendency to stay with the analog techniques.

However, recent advances in digital signal processing have made the digital transmission of voice highly efficient. Until recently digital transmission of speech was possible only by sampling the voice waveform at a sufficiently high rate and then performing an analog-to-digital conversion of each sample. A sufficient number of bits were transmitted for each sample which was sent to reconstruct the waveform at the reciever. The voice waveform must be sampled at aproximately 8,000

samples per second to avoid the loss of clarity. Each of the samples must then be converted to a 6-10 bit number for transmission. The overall data rate using these methods had a lower limit in the neighborhood of 48,000 bits per second.

Recent developments have allowed the voice pattern to be broken down into more basic parameters which are closely associated with the physical production of speech. These parameters vary rather slowly and can be transmitted at a lower rate. Data rates as low as 1200 bits per second have been achieved through the use of these techniques.

These methods are numerical representations of the physical production of speech, and therefore it is easier to alter the characteristics of speech by altering the associated parameters then by trying to alter the waveform directly.

This thesis reviews various digital speech processing techniques for use in a speech modification system. Linear predictive coding (LPC) was chosen for implementation and therefore the theory and practice of this technique are explained in detail. The desired modification of the speech waveform by shifting the poles of its characteristic polynomial, and the regeneration of the altered waveform are discussed and the implementation techniques explained. The IBM 360 computer was used for simulating the techniques developed. This simulation is covered in detail and the computer programs, with results, are provided.

II. SPEECH PRODUCTION AND CHARACTERISTICS

Any digital system for altering speech characteristics must be based on knowledge of those characteristics and the physical structure which determines them.

A. SPEECH CHARACTERISTICS

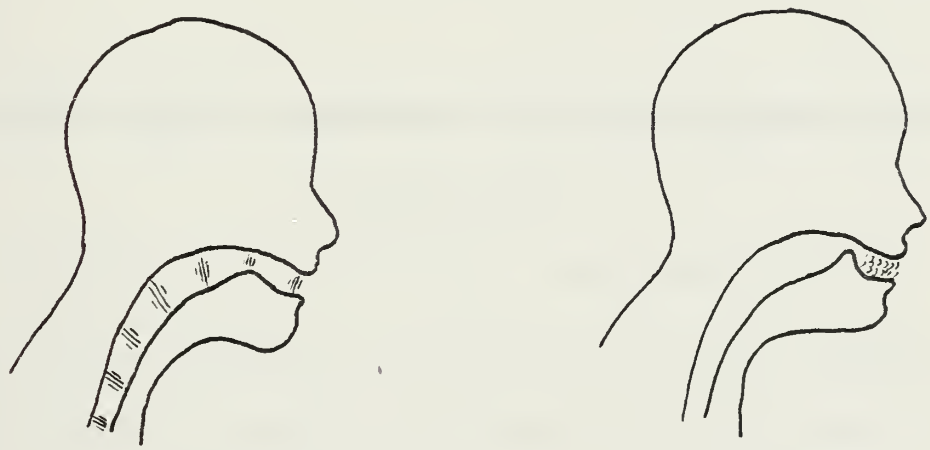
All speech can be broken down into a set of distinctive sounds called phonemes. In the case of American English, there are generally considered to be 42 distinct phonemes which are classified into vowels, diphthongs, semivowels and consonants. Spoken communication is accomplished through various combinations of these sounds and the accurate reproduction of each is a major criteria in judging voice processing systems. Phonemes are generated at a rate of about ten per second. Each phoneme is classified as voiced if vocal cord vibration is the source of the sound or unvoiced if the sound is produced by other means. If the characteristics of a phoneme change from the start to finish, the phoneme is called noncontinuant. Those phonemes which are stationary are called continuant.

The lowest frequency present in a given voiced sound is called the pitch frequency. There are peaks in the spectral representation of a speech sound that are above the pitch frequency which are called formants and are numbered consecutively with increasing frequency. Although two

speakers may produce the same phoneme, the pitch and formant frequencies may be different. However, general relationships may be established between pitch and formant frequencies which are relatively constant from speaker to speaker, producing the same phoneme. If information is to be retained by a speech processing system, it must be able to reproduce at output, the pitch and formant frequency relationship which was present at the input.

B. PHYSICAL SPEECH PRODUCTION STRUCTURE

The vocal tract is a resonant tube with the vocal cords at one end and the lips at the other. The vocal tract acts as a frequency selective filter which has a transfer function that depends on how it is shaped at any given time.



(A) VOICED (B) UNVOICED

FIGURE 1. SOUND PRODUCTION

The input to the vocal tract is caused by either the vibration of the vocal cords at the lower end (figure 1.a) or by the turbulence of air being forced through a

constriction at any of a number of locations along the vocal tract (figure 1.b). The vocal tract acts as a filter with a pulsed input from the vocal cords when producing voiced sounds such as 'a' or 'o'. During sounds caused by the forcing of air through a constriction, fricative sounds like 's' or 'f', the vocal tract acts as a resonant cavity which will have certain characteristic response frequencies. Typical waveforms for voiced and unvoiced sounds are shown in figure 2.

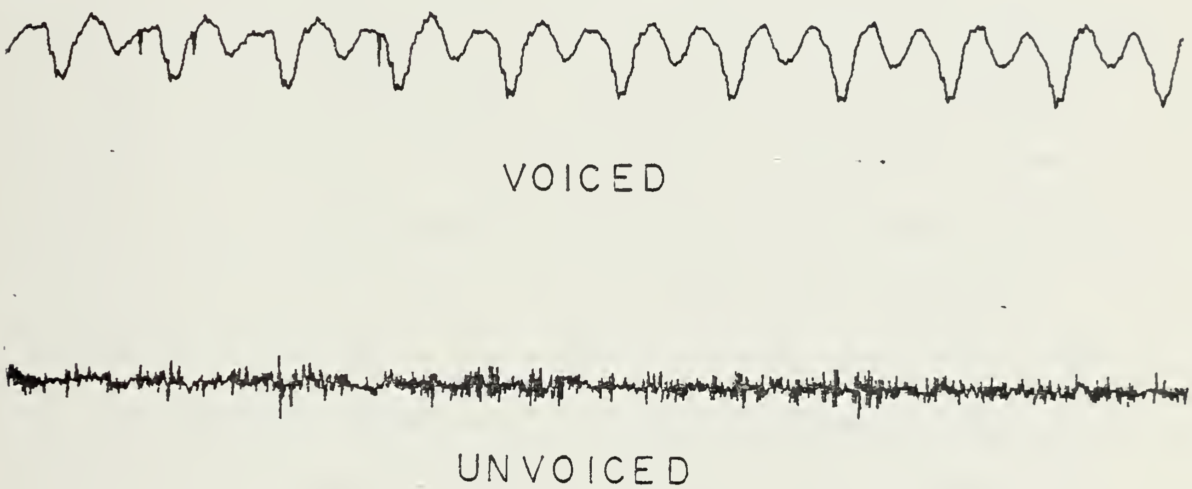


FIGURE 2. TYPICAL WAVEFORMS

Certain characteristics of the vocal tract are changed several times per second to produce different sounds while others such as overall length and the diameter range limits are fixed for a given speaker. A detailed look at each of the types of sounds will insure that the digital processor used has the same flexibility as the actual speaker.

Vowels, voiced continuant sounds, are produced when the vocal cords vibrate causing pulses of air at the bottom

of the vocal tract. The shape of the vocal tract remains fixed during vowel production, acting as a stationary filter to respond to the forcing function.

The production of diphthongs and semivowels is similar to that of vowels except that the shape of the vocal tract is smoothly changed during voicing. Diphthongs and semivowels are noncontinuant, voiced sounds.

The phonemes classified as consonants may actually be further divided into subcatagories of voiced fricatives, unvoiced fricatives, stops and nasals. Fricatives are caused by the steady flow of air through a constriction in the vocal tract which causes turbulant air motion and a seemingly random air pressure pattern. Fricatives are voiced or unvoiced depending on whether the vocal cords are producing pressure pulses at the same time. Stops or plosives are caused by completely closing the vocal tract and then suddenly opening it to quickly start sound production. A stop is classified as voiced or unvoiced depending on the nature of the sound that follows the opening of the vocal tract. Nasals are voiced sounds which are formed when the vocal tract is closed and air is allowed to pass through the nasal cavity. This acts as a feed forward path for the sound and a corresponding change is caused in the total vocal tract response.

C. INFORMATION CONTENT

One of the primary goals of speech processing is the

development of efficient codes for transmitting or storing speech and still allowing it to be reconstructed without excessive loss of information. The source coding theorem states that through the proper choice of coding we can code a source into a bit sequence arbitrarily close in length to the entropy of that source. However, efficient codes are difficult to find for even simple binary sources, let alone a continuous speech source. An estimation of the entropy of a typical speech source provides a useful guage for measuring the data rate performance of any system.

If 'excessive loss of information' occurs only when we don't receive the correct one of the 42 phonemes, the information content of one second of speech is approximately (assuming 10 phonemes are produced per second):

$$H = 10 \sum_{i=1}^{42} P(p_i)\ (-\log P(p_i))$$

where $P(p_i)$ is the probability of the ith phoneme. Assuming further that each phoneme is equally likely,

$$H = 10 \times 42 \times 1/42 \times \log 42 = 54 \text{ bits per second}$$

If the actual probability of each phoneme was used, i.e. they are not equally likely, the value of entropy would be significantly lower.

If 'excessive loss of information' also includes

failure to identify the speaker and failure to indicate the speaker's emotional state the information content is higher. However if we assume that identification of the speaker (one of about two billion) is only reauired once per minute and that the speaker's emotional state (say one of ten) can only change once per second the entropy is still only 58 bits.

$$H(\text{speaker}) = 1/60 \times 10 \quad \times 1/10 \quad \times (-\log(1/10 \ )) = 0.5$$

$$H(\text{emotion}) = 10 \times 1/10 \times (-\log\ (1/10)) = 3.3$$

$$H(\text{phoneme}) = 54 \text{ bits per second}$$

$$H(\text{total}) = 58 \text{ bits per second}$$

Clearly the theoretical limit is not being pushed by the current state of the art in speech coding.

III. DIGITAL SPEECH PROCESSING TECHNIQUES

Digital speech processing techniques may be placed into three general categories based on the assumptions used in their development. The first category is that of waveform techniques where the only primary assumption is that the signal which is being processed is frequency limited to no more than half of the sampling frequency. The second category of spectral methods adds the assumption that the frequency domain characteristics of the speech waveform vary slowly. Finally, the voice tract parameter techniques assume that the physical voice production system can be modeled digitally.

A. WAVEFORM METHODS

Waveform techniques have the characteristic of operating equally well on any low-pass filtered waveform and all are generally based on the familar pulse code modulation. The basic requirements of a waveform quantization method is that the waveform be sampled at greater than twice the highest frequency present and that the samples be quantized into a digital code for transmission. Although this technique is very straight forward, it also requires a high data rate. A waveform sampled 9600 times per second with each sample quantized to 256 levels would require 76,800 bits per second for

transmission. A number of variations (differential modulation and adaptive differential modulation) have been used to reduce the required data rate but have failed to cut the required data rate by more than about half.

B. SPECTRAL TECHNIQUES

1. Short Term Frequency Analysis

These methods deal with the short-term frequency properties of the speech signal. An early spectral method was the channel vocoder. The transmitting processor of the channel vocoder consists of a bank of narrow-band analog filters. The energy passed by each filter is measured and transmitted to the receiver site. It is also determined whether the input speech was voiced or unvoiced and that determination is transmitted. In the receiver an excitation signal, determined by the voicing decision, was fed into a bank of narrow-band filters, each of which had an adjustable gain determined by the received energy measurements.

The same technique can be implemented in an all digital method by replacing the bank of analog filters with digital filters or by performing a discrete Fourier transformation (DFT) on a frame of input samples. The use of the DFT is usually preferred because of computational efficiency and the availability of high-speed DFT array processors. Normally each input frame is windowed to reduce the noise which can be caused by a sharp cut off at

the end of a frame. When this method is used to reduce the data rate required for digital transmission, the total DFT of each frame is not transmitted because the total DFT would require the same number of bits as the frame of samples (assuming both are quantized to the same number of levels). Reduction in the data rate can be accomplished by skipping frames and assuming they are duplicates of the preceeding frame during reconstruction. The number of samples in the frame is also half the number of frequencies resolved by the DFT, therefore the frame length for analysis is choosen as a compromize between accuracy of voice reproduction and the desire for a low data rate.

This method of speech processing would lend itself well to altering the frequency characteristics of voice signals but it requires a relatively high data transmission rate and therefore was not desirable for speech processing in conjunction with place to place communications or with digitally stored speech.

2. Homomorphic Processing

Another method which involves frequency domain processing is homomorphic processing. It is based on the following three principles:

(1) Speech is the convolution of an excitation function and the transfer function of the vocal tract.

(2) Convolution in the time domain is equivalent to multiplication in the frequency domain.

(3) The Fourier transform is a linear transformation, i.e.

$$F(x(t)+y(t)) = F(x(t)) + F(y(t)) = X(w) + Y(w)$$

A method of separating a speech waveform back into these components would help us analyze the speech. Homomorphic processing centers around the efficient deconvolution of these signals.

First the input signal is windowed and transformed via the DFT, to produce the frequency domain representation of the input speech. The time convolution of two signals is equivalent to multiplication in the frequency domain. However knowing the product of two waveforms does little toward gaining knowledge of the multiplicands unless further information is given. The multiplication of the two values at a given frequency is equivalent to adding the logarithms of each. The log is taken of each of the values in the frequency domain representation of the signal which is then equal to the sum of the the log of the frequency domain representation of the excitation function plus the the log of the frequency domain representation of the vocal tract function. However, it is easier to tell the difference between the vocal tract excitation functions in the time domain, so the inverse DFT is taken of the log of the frequency domain function. The function produced is called the cepstrum of the signal. Because taking the inverse DFT is a linear function, and the frequency domain function was the sum of two component functions, the time domain cepstrum must also be the sum of the cepstrum of the

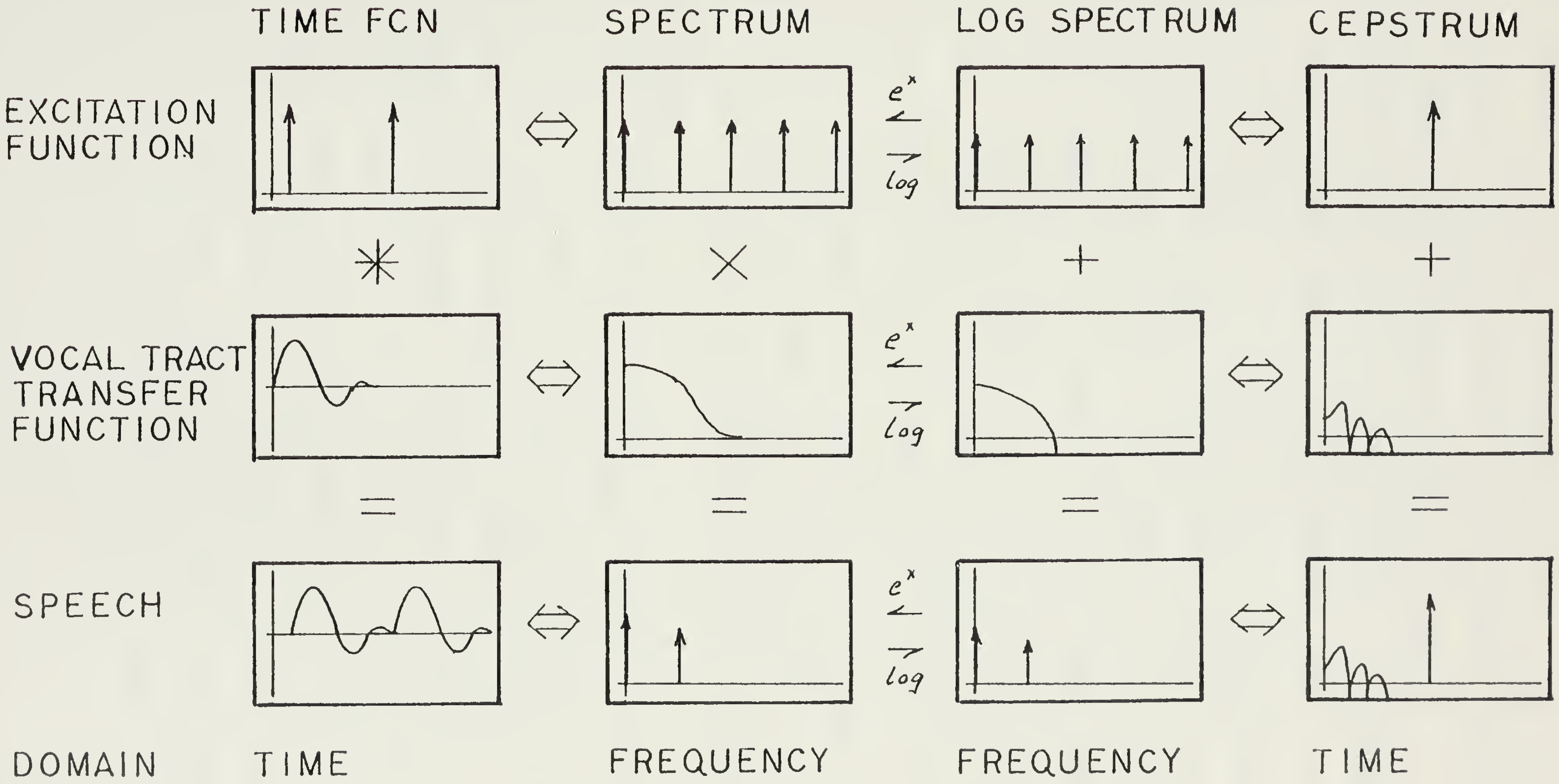


FIGURE 3. HOMOMORPHIC DECONVOLUTION

excitation function and the cepstrum of the vocal tract function. Figure 3 illustrates the relationship between the steps of homomorphic deconvolution of signals.

Examination of the cepstrum between 2.5 and 20 msec. may reveal a peak that is considerably above the background noise level. If a peak is there, the segment is determined to be voiced with the peak occuring at the pitch period. The vocal tract is not long enough to sustain any vibrations for more than 20 msec. after a pulsed input. If there is no peak the segment is considered unvoiced. The cepstrum of the excitation function may be subtracted from the total cepstrum and the remainder considered an estimate of the cepstrum of the vocal tract transfer function. After working backwards to magnitude (vs. log of magnitude) in the frequency domain, the filter coefficients may be determined.

It would be relatively straight forward to alter both the excitation function and the vocal tract transfer function after the total cepstrum is broken into its additive components. However, homomorphic processing was not being widely used for voice communication and this technique was dropped in favor of a more widely used system. As array fast Fourier transform processors become faster and less expensive, homomorphic speech processing may become the dominant speech communication technique.

C. VOICE TRACT PARAMETER TECHNIQUES IN THE TIME DOMAIN

The primary characteristic of this catagory is the close tie between the digital process and the physical structure being modeled. Although homomorphic processing uses the deconvolution of the vocal tract function and the excitation function as a primary tool, the homomorphic process does require transformations to and from the frequency domain and therefore is not included in this catagory. The primary member of this catagory is the linear prediction coding (LPC) process which has shown itself to be among the best and most versitile of the various speech processing techniques.

1. <u>The Speech Model</u>

The speech model assumed and used for LPC is that of a time-varying digital filter which is excited by a wide-band function, either a pulsed input or random noise. This is illustrated in figure 4. The recursive filter used to model the vocal tract is all-pole and has slowly time varying (pseudo-stationary) coefficients. The filter's z-domain transfer function is

$$\frac{Y(z)}{U(z)} = \frac{1}{1-\sum_{i=1}^{p} a_i z^{-i}}$$

or

$$Y(z) = U(z) + \left(\sum_{i=1}^{p} a_i z^{-i}\right) Y(z)$$

or in the discrete time-domain

$$Y(nT) = U(nT) + \sum_{i=1}^{p} a_i Y((n-i)T)$$

From the time domain equation it is clear that the current output Y(nT) is uniquely specified in terms of the current input and the past p output values.

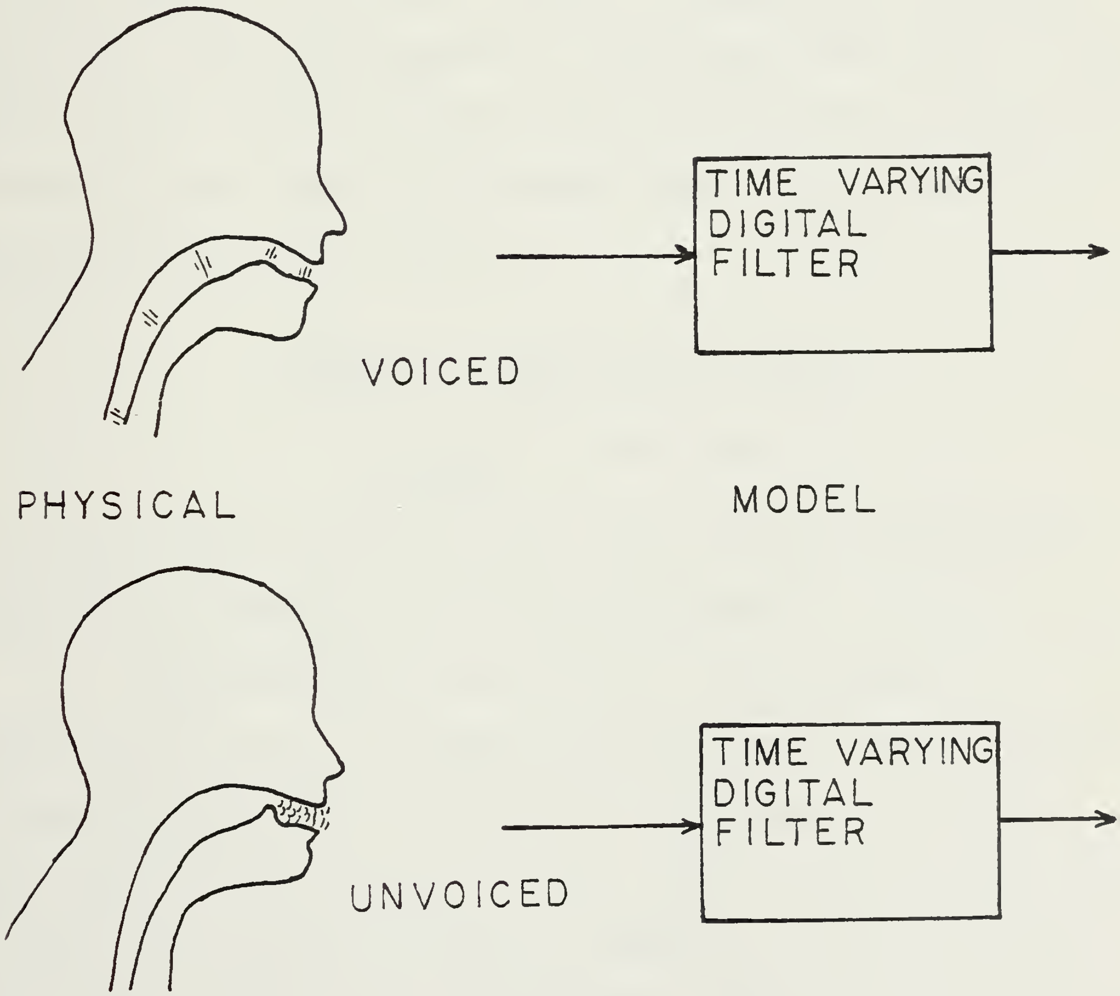


FIGURE 4. SPEECH MODEL

The vocal tract is not always best modeled by an all-pole filter, and particularly nasal sounds would probably be best modeled by a filter which also included zeros. However there is considerable difficulty in rapidly estimating both poles and zeros of a transfer function when only a short segment of the output is available for analysis. However, experience has shown that high quality voice production is possible by using an all-pole filter of adequate order.

The order of the filter required is closely related to the length of the vocal tract. To adequately represent the lower frequency response of the vocal tract, the filter must include recursive delay equal to the delay encountered by sound waves traveling from the vocal cords to the lips and returning to the glottis.

velocity of sound = 344 m/sec
length of vocal tract = 17 cm

$$\frac{2 \times 0.17}{344} = 0.988 \text{ msec}$$

At a sampling rate of 10kHz at least 10 past values would need to be included for an accurate model.

The excitation function for voiced sounds in modeled by a train of pulses at the glottis. Clearly these pulses can not be a perfect set of impulses, but rather must have a finite width and are likely to have a definite shape. Rather than construct a separate filter to change the impulses into the correct shape, additional poles are added to the model so that the combined transfer function

may be calculated at once. Normally two additions poles are adequate for the pulse shape model.

2. <u>Linear Predictive Techniques</u>

Linear predictive analysis is based on the division of speech modeling into modeling of the excitation function and modeling of the vocal tract transfer function. The vocal tract is modeled by computing each sample as a weighted linear combination of previous samples. Linear predictive coding of speech is accomplished by filtering a sampled speech waveform through a filter which is the inverse of the filter which models the vocal tract. If the filter used is the inverse of a good model of the vocal tract, the output will be a good approximation of the excitation function. The various properties of the excitation function, along with the coefficients used in the vocal tract filter are measured and transmitted as shown in figure 5.

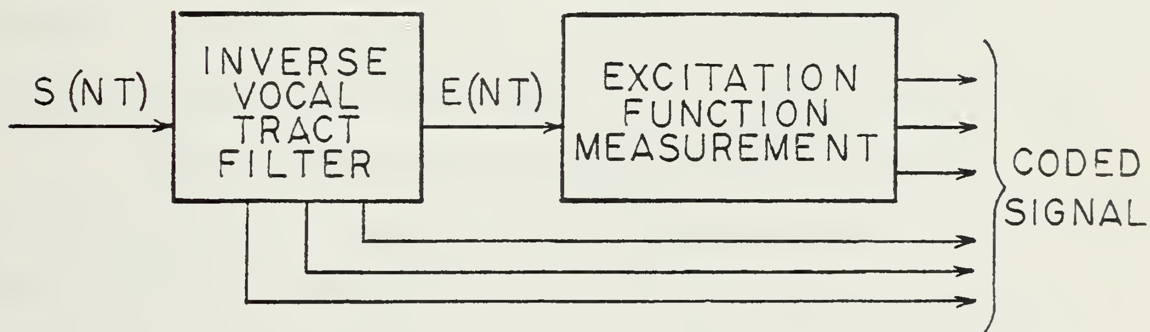


FIGURE 5. ENCODING PROCESS

The received measurements are used in the decoding processor to reconstruct the excitation function and the filter. The process of reconstructing the speech waveform is shown in figure 6.

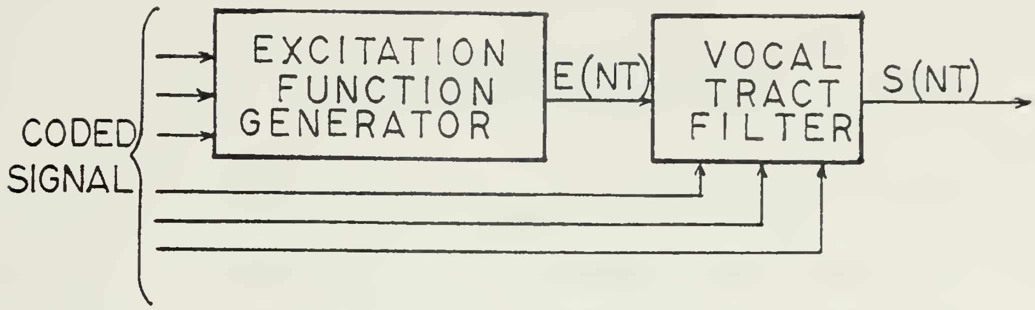


FIGURE 6. DECODING PROCESS

The primary advantage in the use of linear predictive coding of speech is the reduction in the data rate required for transmission or storage. LPC systems have been developed which require data rates from 3000 to 4800 bits per second for high quality voice communication and rates as low as 1200 bits per second have been reported for lower quality but understandable speech production. Highly efficient algorithms have been developed for the encoding and decoding of speech using the LPC technique. When hardware implemented with special purpose, short word length microprocessors, the computations required for two-way communication have been done in 65% of real time.

LPC was chosen as the method to be used for accomplishing the desired voice characteristic modifications. A detailed description of the theory and modeling assumptions follows.

IV. LINEAR PREDICTION THEORY

Linear prediction is an extension of least squares estimation. In the case of one-dimensional linear prediction, it is more commonly labeled as time series analysis when used by statisticians for analysis of everything from population to the stock market.

A. THEORY

It is assumed that each sample of the discrete time series, s(kT), as shown in figure 7 may be approximated by a linear combination of past samples of the time series.

$$s(kT) = \sum_{i=1}^{m} a_i \, s((k-i)T)$$

where s(kT) is the estimated sample value, a_i is the coefficient of the sample i steps past and m is the order of the approximation (and as we will see later the order of the z-domain filter of the model).

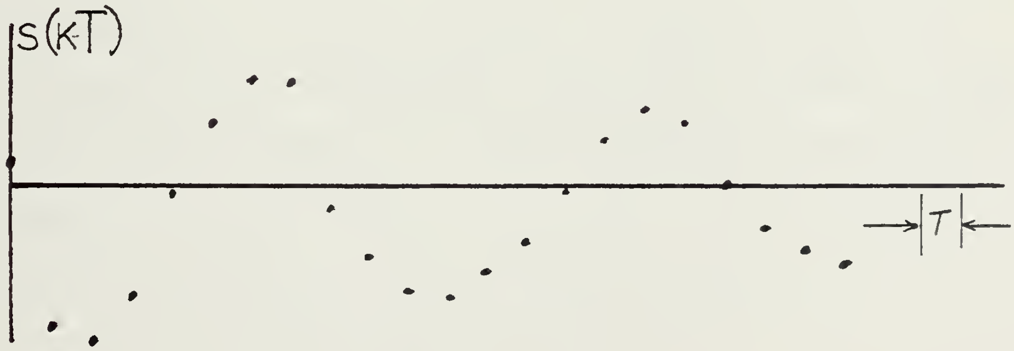


FIGURE 7. DISCRETE TIME SERIES

For a portion of the discrete time series (N samples where $N > m$), a least squares approximation of the weighting coefficients, a_i, may be calculated. The estimate at each point

$$\hat{s}(kT) = \sum_{i=1}^{m} a_i \, s((k-i)T) \qquad 1 \leq k \leq m$$

is subtracted from the actual sample value and the error for each estimate, e(kT) is given.

$$e(kT) = s(kT) - \hat{s}(kT) \qquad 1 \leq k \leq m$$

$$e(kT) = s(kT) - \sum_{i=1}^{m} a_i \, s((k-i)T) \qquad 1 \leq k \leq m$$

To minimize the error (in a least squares sense) the error is squared and summed over all points in the region of interest to obtain an overall error, E.

$$E = \sum_{k=1}^{N} e^2(kT) = \sum_{k=1}^{N} s(kT) - \sum_{i=1}^{m} \left[a_i \, s((k-i)T) \right]^2$$

The derivative of E with respect to each of the coefficients, a_i, is taken and set equal to zero in order to locate the minimum of E. This yields the following m equations.

$$\frac{\partial E}{\partial a_j} = 0 = \sum_{k=1}^{N} \left[2 \left(s(kT) - \sum_{i=1}^{m} a_i s((k-i)T) \right) \frac{\partial}{\partial a_j} \left(s(kT) - \sum_{i=1}^{m} a_i s((k-i)T) \right) \right]$$

$$1 \le j \le m$$

however

$$\frac{\partial}{\partial a_j} \left[s(kT) \right] = 0$$

and

$$\frac{\partial}{\partial a_j} \left[a_i s((k-i)T) \right] = 0 \; , \quad i \neq j$$

$$= s((k-j)T), \; i = j$$

therefore

$$\frac{\partial E}{\partial a_j} = 0 = \sum_{k=1}^{N} 2 \left[s(kT) - \sum_{i=1}^{m} a_i s((k-i)T) \right] (-1) \; s((k-j)T)$$

$$1 \le j \le m$$

removing the constant multiplier

$$0 = \sum_{k=1}^{N} S(kT)s((k-j)T) - \sum_{k=1}^{N} \sum_{i=1}^{m} a_i s((k-i)T)s((k-j)T)$$

$$1 \le j \le m$$

changing the order of summation

$$\sum_{k=1}^{N} s(kT)s((k-j)T) = \sum_{i=1}^{m} a_i \sum_{k=1}^{N} s((k-i)T)s((k-j)T)$$

$$1 \le j \le m$$

Given all of the samples within the summations over N, the above set of m equations in the m unknowns, a_i, can be solved. If only the samples

$$s(kT) \quad 1 \le k \le N$$

are given, the set of equations above can not be solved because of the requirement to know the samples

$$s((1-j)T) \quad 1 \le j \le m$$

However by windowing the samples so that all samples outside the region of interest are zero

$$s(kT) = 0 \quad k \le 0 \text{ and } k > N$$

the summations over N in the set of equations above may be replaced by the autocorrelation of the windowed samples, $s'(kT)$.

$$R(j) = \sum_{k=1}^{N-j} s'(kT)s'((k+j)T) \qquad 0 \le j \le m$$

This assumption may be made because the number of samples, N, is normally much greater than the order, m, of the set of equations. Therefore relatively few samples are lost. The window function used will not significantly alter the samples in the center of the frame, and therefore the resulting coefficients will be a correct approximation for that segment. The set of linear equations may now be written

$$R(j) = \sum_{i=1}^{m} a_i \, R(i-j) \qquad 1 \le j \le m$$

These equations may now be solved for the linear predictive

coefficients, a_i, $1 \leq i \leq m$.

If the system being studied is stationary or we are only considering a pseudo-stationary segment of the system output, and if the order of the model is sufficiently close to the order of the real system, future values of the variable may be calculated recursively from previous values. In the following section we will see how this theory is applied to speech modeling and reconstruction.

B. LINEAR PREDICTIVE CODING FOR VOICE ANALYSIS

The digital model used for speech synthesis is shown in figure 8. The discrete time excitation function is e(nT) and the synthesized speech output is s(nT).

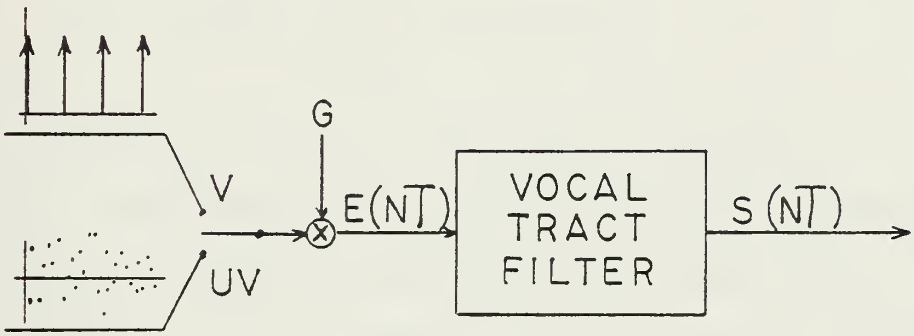


FIGURE 8. SPEECH SYNTHESIS MODEL

The vocal tract filter is assumed to be all-pole and therefore can be represented by the z-domain equation

$$H(z) = \frac{S(z)}{E(z)} = \frac{z^m}{\prod_{i=1}^{m} (z-p_i)}$$

Multiplying out the denominator and dividing both numerator and denominator by z^m yields.

$$H(z) = \frac{S(z)}{E(z)} = \frac{1}{1-\sum_{i=1}^{m} a_i z^{-i}}$$

This z-domain equation is converted to a discrete time domain equation as follows

$$S(z)\left(1-\sum_{i=1}^{m} a_i z^{-i}\right) = E(z)$$

$$S(z) = E(z) + \sum_{i=1}^{m} a_i z^{-i} S(z)$$

$$s(nT) = e(nT) + \sum_{i=1}^{m} a_i s((n-i)T)$$

If the excitation function e(nT) equals zero for a given sample, then this equation is similar to the first equation in the previous section on the theory of linear prediction. The coefficients of the z-domain filter transfer function are equivalent to the linear prediction wieghting coefficients.

Analysis of the sampled speech waveform is used to calculate the prediction coefficients which are then used in an inverse filter to determine the excitation function from the input speech. This inverse filter may be represented as

$$\frac{E(z)}{S(z)} = 1 - \sum_{i=1}^{m} a_i z^{-i}$$

or as

$$E(nT) = S(nT) - \sum_{i=1}^{m} a_i \; s((n-i)T)$$

and is construted as shown in figure 9.

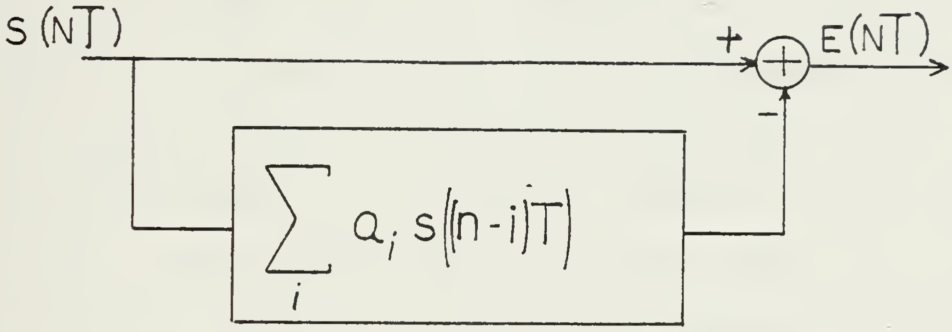


FIGURE 9. INVERSE FILTER

The input speech has been broken into vocal tract characteristics determined by the prediction coefficients and excitation signal characteristics which remain to be determined. During the encoding process the output of the inverse filter may also be considered an error signal because it is the difference between the actual speech sample and the predicted speech sample.

During voiced speech the vocal tract filter in figure 9 acts as a model for the total transfer function which is due to the glottal pulse shape, the actual vocal tract shape and the output reflection at the lips. Idealy during

voiced speech all of these effects are removed by the inverse filter and the error function is a train of impulses at the pitch frequency.

During unvoiced speech the physical excitation function is a pseudo-random air pressure variation caused by turbulence at a constriction somewhere along the vocal tract. This wide-band source is filtered by the portion of the vocal tract between the constriction and the lips. This portion of the vocal tract will resonate at certian characteristic frequencies but normally the number of peaks in the frequency domain response will be fewer than for voiced sounds because of the shorter segment of the vocal tract in use. During encoding of unvoiced speech the output of the inverse filter is pseudo-random because the inverse filter can't predict the output due to the random input.

The speech model is not complete with just the determination of the coefficients of the vocal tract filter. During speech reconstruction it is necessary to know:

(1) Which excitation signal, pulses or noise, to use.

(2) Excitation pulse period for voiced sounds.

(3) The gain multiplication factor.

Although these quantities are not necessarily determined using linear prediction theory, they are none the less required for a working speech encoding/decoding system.

During encoding, the marked difference in the error

signal for voiced and unvoiced speech can be used as the basis for the voiced/unvoiced decision. The energy of the error signal for voiced speech should be rather small in comparison to the energy of the input samples. On the other hand, during unvoiced speech the prediction is poor and most of the energy remains after filtering. The ratio of the average energy or root-mean-square value of the speech samples to the similar quantity of the error signal can be used to make the voiced/unvoiced decision. This ratio is compared to an empirically determined threshold and the segment is considered voiced whenever the ratio is greater than the threshold.

The gain used during reconstruction is the amplitude multiplier of the excitation signal at the input of the vocal tract filter. The gain used during unvoiced speech may be simply the root-mean-square of the error signal. This gain coefficient is multiplied by the output of a random number generator which produces normally distributed numbers with a root-mean-square value of unity.

The gain of voiced speech may also be determined from the root-mean-square value of the error signal. However during reconstruction of voiced speech the entire energy of the excitation signal is concentrated in a series of impulses which should have the same root-mean-square value. The root-mean-square value of a series of discrete-time impulses with amplitude, a, and a period, p, intervals is approximated by

$$rms = \left[\frac{1}{N}\sum_{i=1}^{N} x_i^2\right]^{1/2}$$

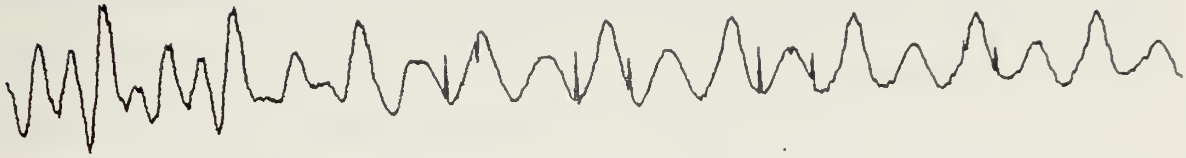
$$rms \cong \left[\frac{1}{N}\,\frac{N}{P}\,a^2\right]^{1/2} \qquad N \gg p$$

$$rms \cong a\,p^{-1/2}$$

The output of a unit impulse generator should then be multiplied by

$$G = rms\;p^{1/2}$$

to insure that the same energy is input to the vocal tract filter as was output by the filter during encoding. The above method for calculating the gain needed during reconstruction is based on the assumption that the prediction error for voiced speech is caused entirely by the physical excitation function of the speaker. However the prediction error may be increased because the vocal tract was changing shape rapidly during the analysis frame or because of background noise at the microphone which would not be removed by the inverse filter. Either of these would cause an unwanted gain increase during reconstruction. A typical voiced speech waveform and the error signal generated from it are shown in figure 10.



(A) VOICED SPEECH WAVEFORM

(B) ERROR SIGNAL WAVEFORM

FIGURE 10.

The reliable determination of the pitch period of voiced speech is a problem for which the ideal solution is still undetermined. The periodic increase in the amplitude of the error signal at the pitch period is shown in figure 10(b) and suggests the use of the error signal in pitch period determination. A number of algorithms exist for determination of the pitch period which generally involve various combinations of the following processes.

(1) Raising the error signal to a given power.

(2) Low-pass filtering of the error signal.

(3) Windowing the error signal.

(4) Calculating the autocorrelation function of the filtered error signal.

(5) Picking the peaks of the autocorrelation function.

Experience has shown that pitch determination is computationally as difficult as the LPC parameter

determination and the literature on the subject illustrates the trade-off between hardware, software, computation time and reliability from method to method.

C. LPC COMMUNICATION SYSTEMS

A review of existing LPC communication hardware is useful because any method which alters formant and pitch characteristics of speech will be most successful if it is compatable with these systems.

Currently off-the-shelf microprocessors are not fast enough to handle the algorithms described in real-time. However special purpose units which are designed along computer lines, do meet the real-time criteria. On the surface the word 'computer' might not seem to fit these special purpose machines, but a closer look will reveal that each has components which are the same as those of a computer: stored programming, memory, input, output, an arithmetic logic unit (ALU), an instruction set, and control components. Two processors which were developed at MIT's Lincoln Laboratory will be used to illustrate the state of the art in LPC voice terminals and certain similarities in their architecture will be evident. The first processor is the more flexible of the two and is designed to handle a wider varity of algorithms. The second was developed about a year later and was designed specifically for LPC algorithms with only minor changes.

The first processor to be covered is the Lincoln

Digital Voice Terminal (LDVT) which was designed and constructed at the Lincoln Laboratory during the 1973-75 time frame. This processor is capable of carrying out 18 million basic instructions per second with a 16-bit by 16-bit multiplication taking four times as long. The execution time for each instruction is 165 nsec. which seems to conflict with the instruction rate. This is resolved by the pipelining of the three portions of each basic instruction: fetch, decode, and execute. The processor has separate memories for data and the program. The data memory capacity is 512 16-bit words and the program memory contains 1024 16-bit instructions. The pipeline instruction processing requires that the buses to and from the ALU be seperate and each is unidirectional.

Figure 11 shows the data paths of the LDVT (none of the control or timing lines are shown). There are four active registers: the P register which is the program counter with multiplexed inputs from the address portion of the instruction, the ALU, the sum of the X register and the address portion of the instruction, and itself incremented by one; the X register which is used for indexing memory addresses; the A register which is the accumulator; and the B register which is actually a pair of registers used for input and output.

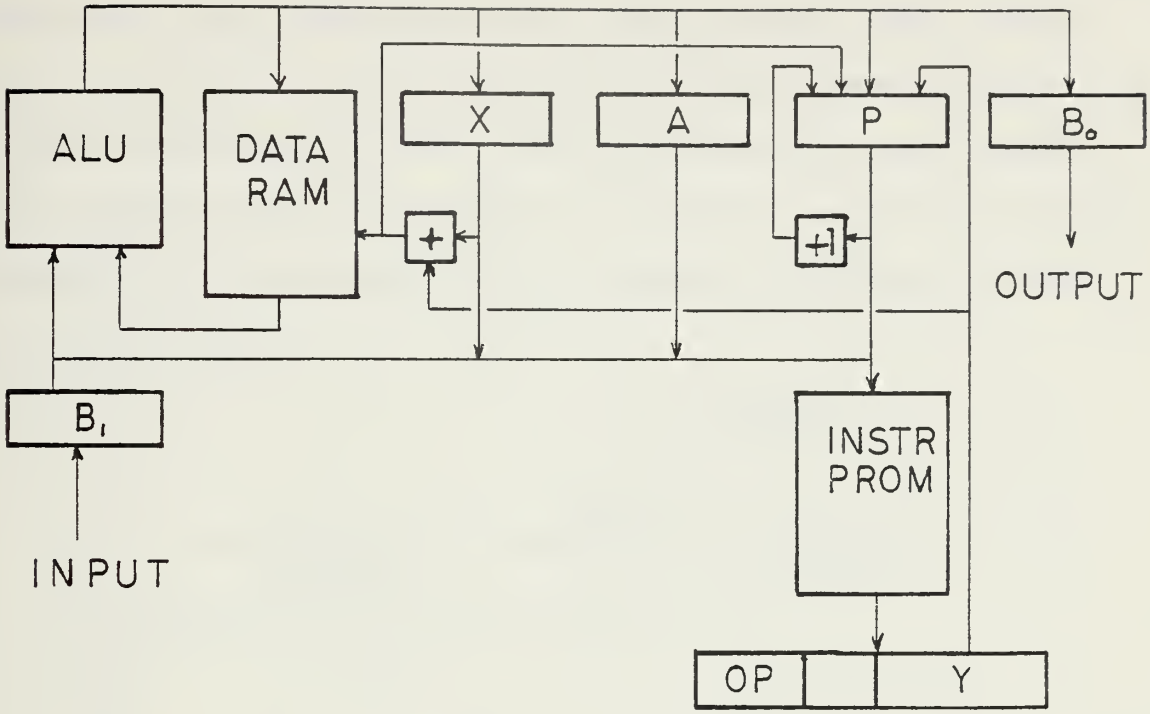


FIGURE 11. LDVT DATA FLOW

The ALU of the LDVT as shown separately in figure 12, has two sections: a standard programmable ALU which performs logical, addition and compare operations; and a 16-bit by 16-bit multiplier array which provides a 32-bit result in just 4 cycles. Either of these may be used with any input, however due to their common input and output only one may be used at a time.

It is significant to note some of the requirements brought on by the pipelining of the instructions. The device does not have a main bus over which data flows in both directions. Generally all data flow is unidirectional and in the case of the ALU input buffer registers are

needed to hold the data for the instruction being executed while the next instruction may have already read a value from memory and put this on the ALU input line. In addition to LPC algorithms at 2400, 3600 and 4800 bits per second, the LDVT has been programmed for adaptive predictive coding at 8000 bits per second and as a channel vocoder at 2400 bits per second.

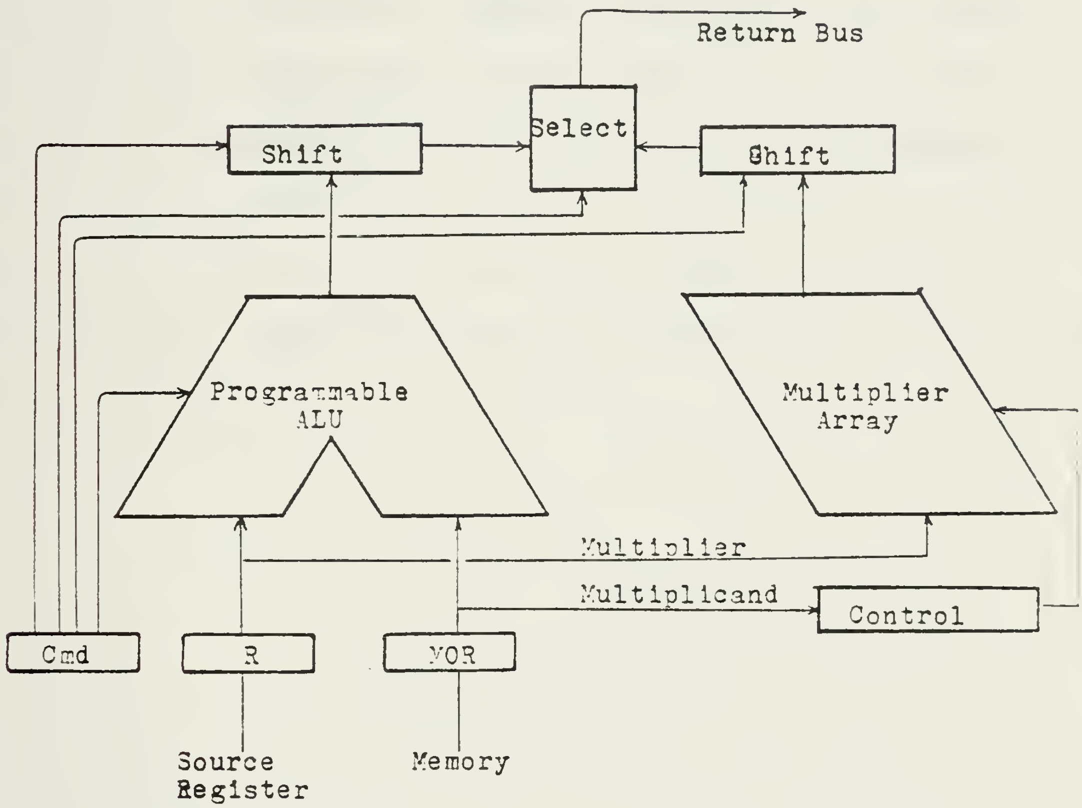


FIGURE 12. LDVT ALU

The second speech processor is the Linear Predictive Coding Microprocessor (LPCM) which is disigned strictly as a low cost LPC terminal. The basic cycle time for this machine is 150 nsec. The data memory has 2K 16-bit words of which 1.5K is ROM and 0.5K is RAM. The program memory contains 1K of 48-bit words. The LPCM is almost free of

instruction decoding, with the only exception being the ALU operation. Figure 13 shows the instruction format and in figure 14 it is evident that parts of the instruction register are being input as control functions. Figure 15 is a block diagram of the LPCM and shows the two buses and the large number of registers needed to control the data flow.

While these machines have varying degrees of adaptability, it does not appear that either could handle the additional computations described in the following sections without major hardware modifications. However, a special purpose LPC code converter which could be used in conjunction with an existing terminal could probably be developed which would operate in real-time and not load the existing processor.

Bits	Field	Description
11	F	Constant or address index for use by the CPE
1	HLT	Halt when set
1	ECY	Use saved carry bit during this instr
1	SCY	Save carry bit from output of CPE
1	SIL	Set interrupt lockout
1	RIL	Release interrupt lockout
1	TST	Control of jump instructions and tests
1	R	
1	S	
3	JPC	
4	B	Supply address for CPE B latch
2	U	Not used
4	A	Supply address for CPE A latch
3	OC	Strobe selected register to receive output of CPE
3	IC	Select CPE input line
3	I_d	Control routing of ALU output in CPE
3	I_s	Instruction for ALU
3	I_o	
1	C_o	

FIGURE 13. LPCM INSTRUCTION FORMAT

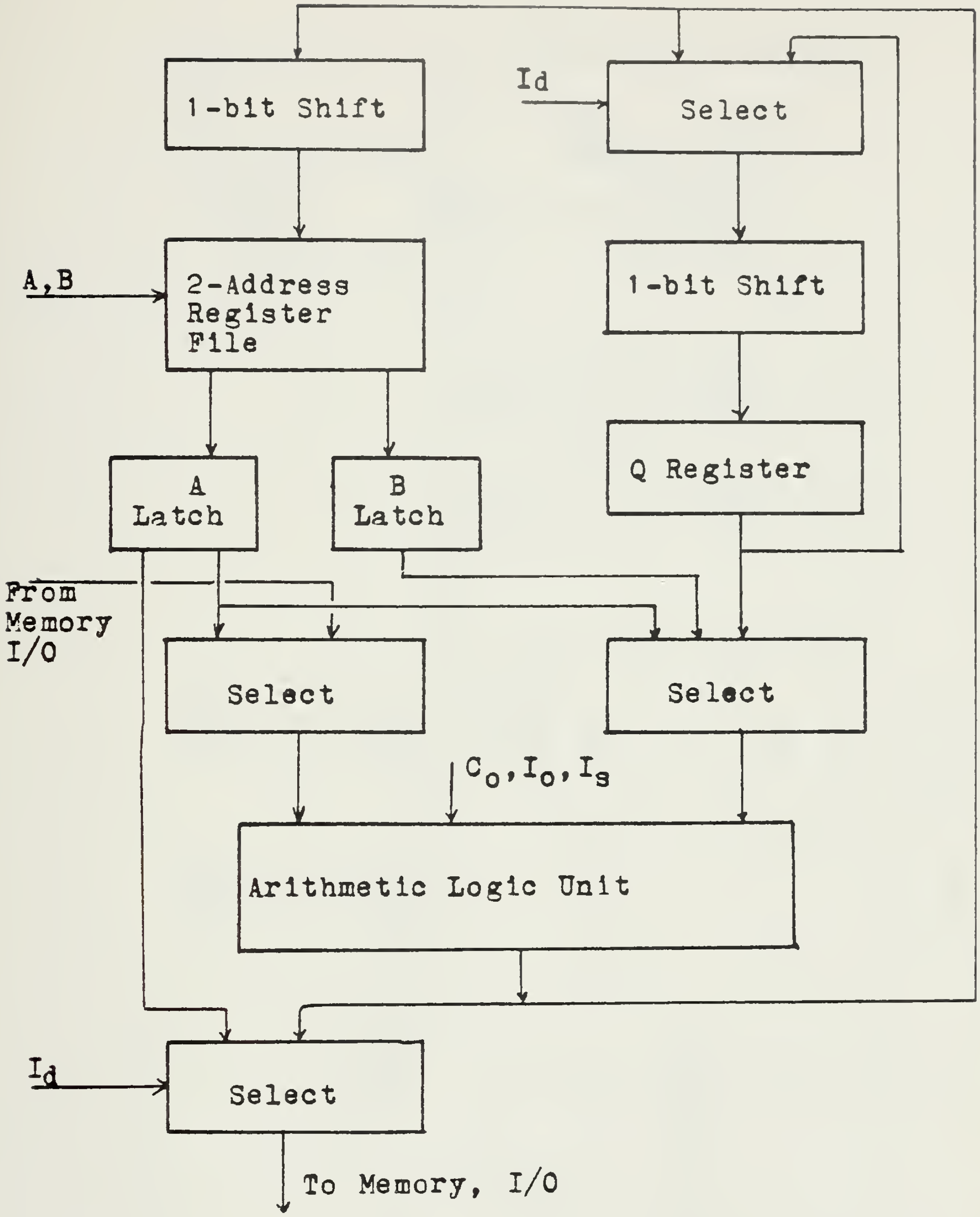


FIGURE 14. LPCM CENTRAL PROCESSOR

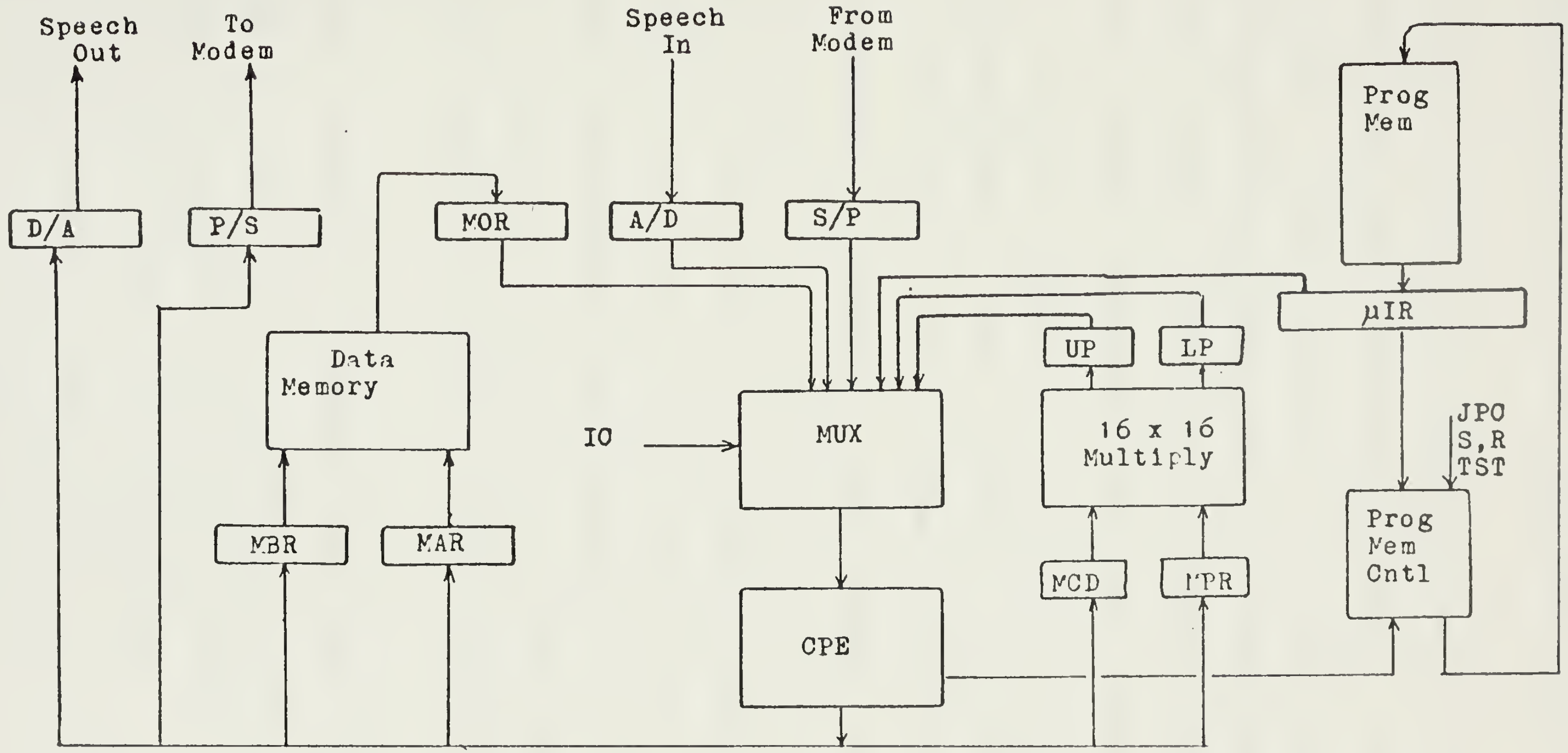


FIGURE 15. LPCM BLOCK DIAGRAM

V. ADJUSTMENT OF VOCAL TRACT PARAMETERS USING LPC

One reference to voice characteristic modification was found by the author [Atal and Hauneur, 1971]. Although scaling of pitch, formant frequency and formant bandwidth was stated to have been accomplished, no description of the work was given. Other literature did provide useful information on formant frequencies and pitch periods which are typical for various speakers. It should be noted that there is a considerably larger variation, from speaker to speaker, in pitch period than in formant frequencies. As an example, two speakers, saying the same phoneme could easily have pitch periods that varied by a factor of two, yet have only a 10-20 per cent variation in formant frequencies. Different physical structure (vocal cords and the vocal tract) produce these speech characteristics (pitch period and formant frequencies, respectively) and therefore their variation from speaker to speaker is only partially correlated.

The coded information produced from input voice by the LPC processor is very closely related to the physical structure that is producing the sound. On output, speech is reconstructed from the gain, pitch period and voice/unvoiced parameters as well as the vocal tract prediction coefficients. The gain and pitch period can be varied as they stand but the variation of the prediction

coefficients is somewhat more complicated. The goal of varying these coefficients before reconstruction is to have the output voice have different pitch period and formant frequencies while retaining a natural sound and retaining the same information, i.e. the same sequence of phonemes and voice inflection.

Voice characteristics are associated with certain parameters of the LPC code. First, formant frequencies and bandwidths are associated with the LPC coefficients. The amplitude of the output voice is associated with both the gain coefficient and the formant bandwidths. The relationship between output amplitude and the formant bandwidth is due to the increased energy in the impulse response of a narrow bandwidth (high Q) transfer function. This is noted physically by the fact that speakers with highly resonant voices may speak louder for the same amount of energy expended. The pitch period is controled by the pitch period coefficient only. Finally, the voice/unvoiced decision would normally not be changed. The exception would be if one was reconstructing whispered speech (the vocal cords are stationary) from normal speech.

A. ADJUSTMENT OF FORMANT FREQUENCY AND BANDWIDTH

The vocal tract model we are using has all real coefficients in the z-domain polynomial. Following directly from this is the fact that all poles must fall either on the real axis of the z-plane or in complex conjugate pairs.

Each of the complex conjugate pairs is associated with one formant (resonator) of the speech model. The vocal tract transfer function is the product of these resonator transfer functions which are each of the following form

$$H_f(z) = \frac{1}{1-2e^{-2\pi(BW)T_s}\cos(2\pi F\, T_s)z^{-1} + e^{-4\pi(BW)T_s}z^{-2}}$$

where F is the center frequency of the formant, f , and BW is the bandwidth of the formant. The pole locations associated with this transfer function are

$$z = x \pm jy$$

This pair of poles must be moved in order to alter the frequency and bandwidth of this resonant section of the vocal tract model, but this must be done carefully so that the poles remain inside the z-plane unit circle. If the desired modification of the input speech is to reduce the bandwidth (increase Q) of the formants, the poles must be moved closer to the unit circle. If the distance from the center is multiplied by a constant factor, there is a danger of moving poles outside the unit circle and thereby causing instability during reconstruction. However, the magnitude of the pole is always less than one and may be raised to any positive power without danger of crossing the unit circle. It is shown as follows that raising the magnitude to a factor is equivalent to multiplying the formant bandwidth by that same factor.

The transfer function with the complex conjugate poles above is:

$$H(z) = \frac{1}{1-2x z^{-1} + (x^2+y^2) z^{-2}}$$

However with the pole locations in polar form

$$x = A \cos \theta \qquad Y = A \sin \theta$$

and making use of

$$\cos^2 \theta + \sin^2 \theta = 1$$

the equations becomes

$$H'(z) = \frac{1}{1-2A \cos\theta\ z^{-1} + A^2 z^{-2}}$$

Setting the terms of the characteristic equations equal we get

$$2A \cos\theta = 2e^{-2\pi (BW) T_s} \cos(2\pi F T_s)$$

and

$$A^2 = e^{-4\pi (BW) T_s}$$

when solved for A and θ give

$$A = e^{-2\pi (BW) T_s}$$

$$\theta = 2\pi F T_s$$

and inversely

$$F = \theta / 2\pi T_s$$

$$BW = (-\ln A) / 2\pi T_s$$

If new formant characteristics, F' and BW', are desired where

$$F' = \gamma F$$

and

$$BW' = \alpha BW$$

they may be implemented by moving the poles of the characteristic equation so that

$$\theta' = \gamma\theta$$

and

$$\ln A' = \alpha \ln A$$

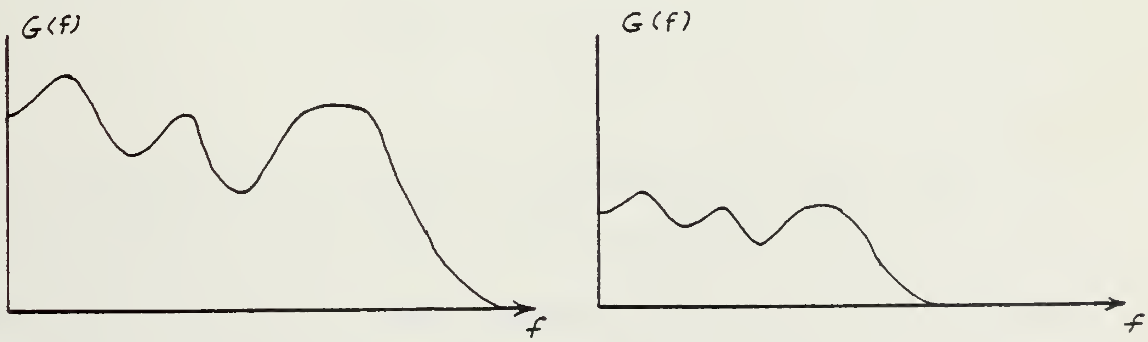
which reduced to

$$A' = A^{\alpha}$$

This method of implementing the pole shifts guarantees that no unstable poles will be created and is used in the following section in the realization of a LPC voice modification system.

B. GAIN ADJUSTMENT

The filter coefficients reconstructed from the relocated poles above may not have the same zero frequency gain characteristic as the filter used for inverse filtering during encoding. This situation can be illustrated graphically by the two vocal tract transmission characteristics shown in figure 16.



BEFORE PROCESSING (A) AFTER PROCESSING (B)

FIGURE 16. FORMANT GAIN

Although the formant frequencies in 16(b) are lower than the corresponding frequencies in 16(a) as was desired, the overall gain was also changed. This would cause the reconstructed speech to be much softer than desired.

A solution to this problem was to adjust the excitation function gain used during reconstruction. This adjustment factor would be equal to the ratio of the zero frequency gains of the original and modified vocal tract filters. The vocal tract has the following z-domain transfer function.

$$H(z) = \frac{1}{1+\sum_{i=1}^{p} a_i z^{-i}}$$

The above equation can be evaluated at

$$z = e^{j\pi f_i/f_s}$$

to obtain the gain at frequency f_i. Evaluating the above transfer function at f=0 yields the following equations.

$$z^{-i} = 1$$

and

$$G(0) = \frac{1}{1+\sum_{i=1}^{p} a_i}$$

This equation can be easily evaluated for both the coefficients of the vocal tract transfer function calculated from the input sequence and the coefficients calculated from the altered pole locations. The gain multiplication factor is then multiplied by the energy

measured in the error signal to get the excitation gain to be used during reconstruction.

C. PITCH PERIOD ADJUSTMENT

The adjustment of the measured pitch period may almost go without explanation except to note that if the pitch period is increased and all other coefficients remain unchanged, the output speech would be softer. This is due to the reduced energy (impulses less often) being input to the vocal tract filter and the resulting lower energy in the output speech.

VI. COMPUTER SIMULATION OF PITCH AND FORMANT MODIFICATION

The process of pitch and formant modification was carried out on the IBM 360 computer with the input and output being accomplished on a hybrid system consisting of a COMCOR 5000 analog computer and an XDS 9300 digital computer. The interface between the XDS 9300 and the IBM 360 was seven track digital magnetic tape. All work was done on five second segments to allow sufficient length for analysis while not using excessive computer processing time.

A. VOICE INPUT AND DIGITAL SAMPLING

The input voice was recorded on a standard single tract audio tape recorder at 7 1/2 inches per second (ips). Recording was done with a high quality microphone in a quiet but not sound-proof room. This digitizing was done at half speed to allow the digital computer to write the data onto tape without missing any data. This recording was played back at 3 3/4 ips with the output directed to an amplifier of the analog computer. The voice was amplified to a level appropriate for the analog computer (a $\pm$100 volt machine). The amplifier output was passed through two forth-order analog filters set at 2350 Hz and 2400 Hz cut off frequencies. The output of the filters was then put into a sample and hold circuit at the input of a 14-bit

analog to digital converter. The 14 bits produced were read by the XDS 9300 and placed in the most significant bits of the 24 bit XDS 9300 computer word. This process is illustrated in figure 17.

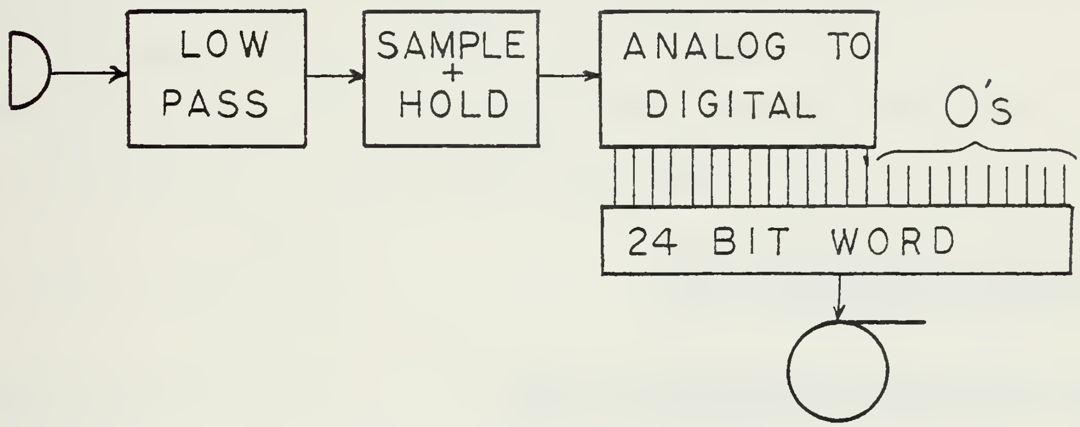


FIGURE 17. DATA ACQUISITION

The sampling rate used was 5000 Hz. However the voice recording was played back at half speed and therefore the equivalent lowpass filter cut off and the equivalent sampling rate were about 4750 and 10,000 Hz respectively.

B. XDS 9300 OPERATION

The operation of the XDS 9300 during the input phase was simply to read the data available at the output of the analog to digital converter and place this data in an array. When an array of 1024 samples was filled it was written onto a seven track magnetic tape. This was done continuously so that no data was lost between blocks. The voice segment as it existed on the seven track tape consisted of 50 blocks of 1024 samples. Each sample was

recorded in a integer format ranging from +8388607 to -8388607 ($\pm(2^{**}23)-1$). This tape was then used as the input to the IBM 360.

C. IBM 360 INPUT PREPARATION

When the 24-bit word, seven track tape created by the XDS 9300 was read by the IBM 360, the machine representation of the values was not correct. This was due to the addition of the eight bits shown in figure 18.

24-Bit XDS 9300 Word

32-Bit Word Read by IBM 360

Corrected IBM 360 Word

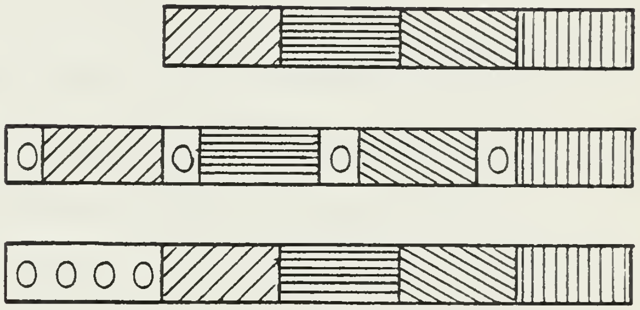


FIGURE 18.

The data conversion program (Appendix A.1) was used to read the data from the seven track tape and move the bits of each value as required. The program did not make the conversion from ones complement representation (XDS 9300) to twos complement reoresentation (IBM 360) because any error caused would be well below the 14-bit quantization error. At this point the data was converted to floating point representation with values between ± 100.0 and the average value of each sequence was calculated and subtracted from each data point. This insured that the input was a zero mean function. Each data sequence was

written into a separate file of a standard nine track IBM 360 tape for ease of further handling.

D. SCOPE OF SIMULATION PROGRAM

The goal of this research was to demonstrate the feasibility of voice modification and as a result only certain areas were studied. Specifically, all programming was done with the standard IBM 360 floating-point arithmetic, making no allowance for the effects which would be caused by the shorter word length and integer representation used in most voice processing systems. Further study of that area is warranted and would be especially critical in the determination of the pole location, which is covered later.

The system degradation by background noise in the input speech was not studied except to note that the voiced/unvoiced deciion threshold would need to be adjusted for a noise environment.

Although the programs were written to allow variation in the order of the prediction, number of samples per frame and sampling interval, these were not varied. A 12th order voice tract filter was used throughout and proved to be satisfactory. The analysis frame length was 25.6 msec. (256 samples) and also remained unchanged. In any future use of these programs with a different frame length, attention would be required by the input format to insure that the analysis frame length is an integral multiple of

the input record length.

Finally, in the following description of the programs the term 'LPC coefficients' will refer to the coefficients of the vocal tract model filter. The term 'LPC parameters' will refer to the entire set of parameters needed to reconstruct the output speech, i.e. the LPC parameters consist of the LPC coefficients, the gain parameter, the pitch period and the voicing indicator.

E. LPC ENCODING

The first step of the encoding process was to determine the filter coefficients. These coefficients were used in the inverse filter for determination of the error signal. The root mean square values of the input and error signals were compared to determine if the frame was voiced or unvoiced. Finally the pitch period was determined for voiced frames. This program is listed in Appendix A.2.

1. LPC Coefficient Determination

Determination of the LPC coefficients was done with the autocorrelation method in the subroutine named AUTO. First, the input data, $s(n)$, was windowed by one of four available windows producing a temporary array, $t(n)$, of the windowed data.

$$t(n) = W(n) \times s(n)$$

The discrete autocorrelation of the temporary array was calculated for the discrete displacements of zero to the predictor order, p.

$$R(j) = \sum_{i=1}^{N-j} t(i)\ t(i+j)$$

$$0 \le j \le p$$

The next step was the solution of the following matrix equation.

$$\sum_{j=1}^{p} R(|i-j|)\ a_j = R(i)$$

$$1 \le i \le p$$

The auto correlation matrix in always positive definate, symetric and all values along a given diagonal are equal. A particularly efficient method of solution is available. This method is attributed to Durbin |Makhoul, 1975| and is implemented in subroutine COEFF. Durbin's algorithm is recursive and calculates the predictor coefficients for the Kth order from the coefficients for the (k-1)th order. The jth coefficient for the kth order predictor is $a_j(k)$. The recursion formulas follow.

$$E(0) = R(0)$$

$$a_j(k) = \left[R(j) - \sum_{i=1}^{j-1} a_i(j-1)\ R(j-i)\right] / E(k-1)$$

$$1 \le j \le p$$

$$a_j(k) = a_j(k-1) - a_k(k)\ a_{k-j}(k-1)$$

$$1 \le j \le (k-1)$$

$$E(k) = (1-a_k(k)^2)\,E(k-1)$$

E(k) is the prediction order error resulting from limiting the predictor order to k.

During the programming of COEFF the subroutine TEST was written to perform and print the results of the matrix multiplication. During the initial testing of the program various window functions were used in AUTO, however the prediction order error did not change significantly with the window function used.

Certain researchers have noted that a lower order filter may be used during unvoiced speech. If this is desired, the coefficients for the lower order filters could be stored during the recursive steps of the algorithm above and later, when the frame is determined to be unvoiced, the lower order filter coefficients would be available without further calculation.

The coefficients, a_i, used in the main program are the coefficients of the characteristic polynomial of the filter with a assumed to be unity.

$$H(z) = \frac{1}{\sum_{i=0}^{p} a_i z^{-i}}$$

Therefore the negitive of the values calculated in COEFF were returned to the main program.

2. <u>Error Signal Determination</u>

The error signal, e(n), is determined by subtracting the predicted sample value, $\hat{s}(n)$ from the actual value, s(n).

$$e(n) = s(n) - \hat{s}(n)$$

$$s(n) = -\sum_{i=1}^{p} a_i \; s(n-i)$$

$$e(n) = s(n) + \sum_{i=1}^{p} a_i \; S(n-i)$$

This operation is carried out by subroutine ERR. In order to make a correct error determination at the begining of each frame, a number of samples equal to the order of the predictor were saved from the end of the previous frame. This eliminated additional error signal energy caused by poor begining of frame prediction and reduced the possibility of an incorrect voicing decision. Another possible solution to this problem would be just not analyzing the error for the first few samples of each frame and making the appropriate changes in the following routines that use the error signal.

3. <u>Voicing Decision</u>

A comparison of input signal energy and the error signal energy was used to determine if a particular frame is voiced or unvoiced. Although the root mean square value of each set of data is actually proportional to the square

root of the energy in the signal, the root mean square value was used in this comparison. Whenever the root mean square value of the input signal divided by the root mean square value of the error signal was greater than a threshold value, the frame was determined to be voiced and the voicing indicator was set to one. Otherwise the voicing indicator was set to zero.

4. Pitch Period Determination

The error signal was used in subroutine PITCH for determination of the pitch period of each voiced frame. First the error signal was passed through a recursive 5th order Butterworth filter with an 800Hz cut off, to smooth the signal. Extra samples of the error signal and filtered error signal were saved from frame to frame (zeroed during unvoiced frames) to insure a correct filtered error signal at the begining of each frame. The degradation of the system if this was not done was negligible but plots of the filtered error signal would have shown discontinuities at the begining of each frame if this had not been done. The frame was windowed to eliminate end effects and the autocorrelation function of the filtered error signal is calculated. The portion of the autocorrelation function from 12 to 180 samples was searched for peak values and the pitch period set equal to the location of this peak. Figure 19 shows a typical autocorrelation function and the portion of the curve searched for the peak value. The peak picking algorithm checked to insure that the value chosen

was not on the downslope of the center peak and was not a minor peak with a larger peak at a longer pitch period.

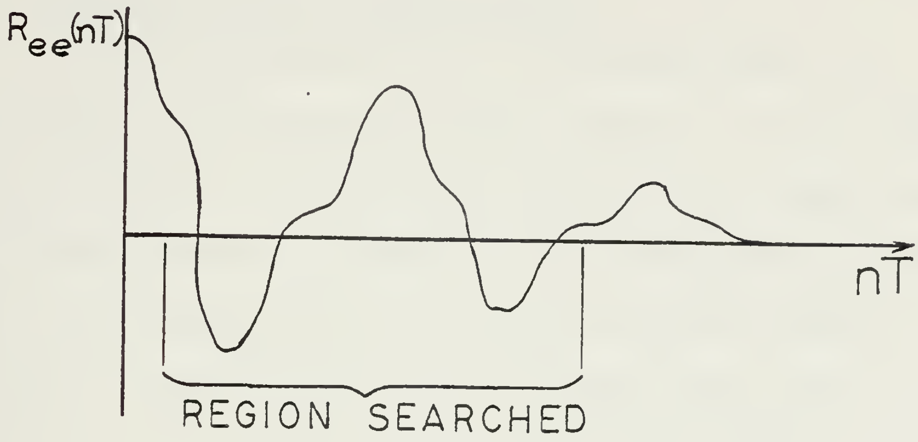


FIGURE 19.

Although this pitch determination algorithm worked satisfactorily in this program it is probably not as accurate and flexible as certain other, more complicated techniques available. It was used only for pitch periods from about 3 to 9 msec., but was satisfactory for them.

F. LPC PARAMETER MODIFICATION

The purpose of the program was to demonstrate the modification of voice characteristics. The system was designed so that only the LPC parameters were needed to make the desired modifications. No other measurements of the input speech are needed. Of the parameters calculated from the input speech, only the voicing indicator remained unchanged. The LPC coefficients are varied as required by the desired formant frequency and bandwidth changes require. The pitch period is varied separately and the gain

is adjusted to correct for changes caused by formant bandwidth modification.

1. LPC Coefficient Modification

The modification of the LPC coefficients is accomplished by three subroutines: POLES, ALT, and NEWCF. Subroutine POLES calculates the z-plane pole locations from the LPC coefficients. Subroutine ALT changes the locations of the poles according to the various scale factors specified by the main program. The new predictor coefficients are calculated by subroutine NEWCF.

The predictor coefficients, a_i, are provided to subroutine POLES to get the p order z-domain polynomial which is factored into its component roots, the z-plane poles of the vocal tract filter. This factorization is done with library routine ZRPOLY which was sufficiently accurate and produced complex conjugate pairs which were exact complex conjugates. This simplified the problem which came up later, of separating the real poles and the complex conjugate pairs so that the proper scaling factor could be applied to each. The input polynomial had all real coefficients and therefore all the roots are real of in complex conjugate pairs. These poles are placed in a complex array and returned to the main program.

The subroutine ALT was provided with the complex array of pole locations and it separated them into separate arrays of real and complex poles. Each complex conjugate pole pair was entered as one entry in the complex pole

array. The scaling factors provided to subroutine ALT consisted of:

(1) FSC - Formant frequency scaling factor

(2) BSC - Formant bandwidth scaling factor

(3) RSC - Real pole scaling factor

(4) RLIM - Real pole magnitude limit

(5) SP - Sampling period

The polar coordinates were determined for each pair of complex conjugate poles and the magnitude, A, and angle, θ, of each were considered separately. The magnitude was raised to the power of the bandwidth scale factor and the angle was multiplied by the frequency scale factor.

$$A' = A^{BSC}$$

$$\theta' = \theta \times FSC$$

The modified magnitude, A', and angle, θ', were used to determine the complex location and the calculated pole and its conjugate were put in the pole vector for output. During the alteration process each complex pair of poles was checked against a constant magnitude of 0.98 to insure that numerical instability or repeated impulses would not cause excessively large outputs.

Each real pole was multiplied by the real pole scale factor and checked to insure that the magnitude was less than the limit prescribed. The effects of varying the real poles was not studied and a real pole limit of 0.95 proved to guarantee sufficient damping of the output to

provide a nearly zero mean output.

The poles from both the real and complex pole arrays were combined into one array for return to the main program. Subroutine ALT also provided graphical and printed output of the pole locations, before and after modification when this was desired. Figure 20 is an example of the graphical output which shows the z-plane pole locations before and after modification, in relation to the unit circle.

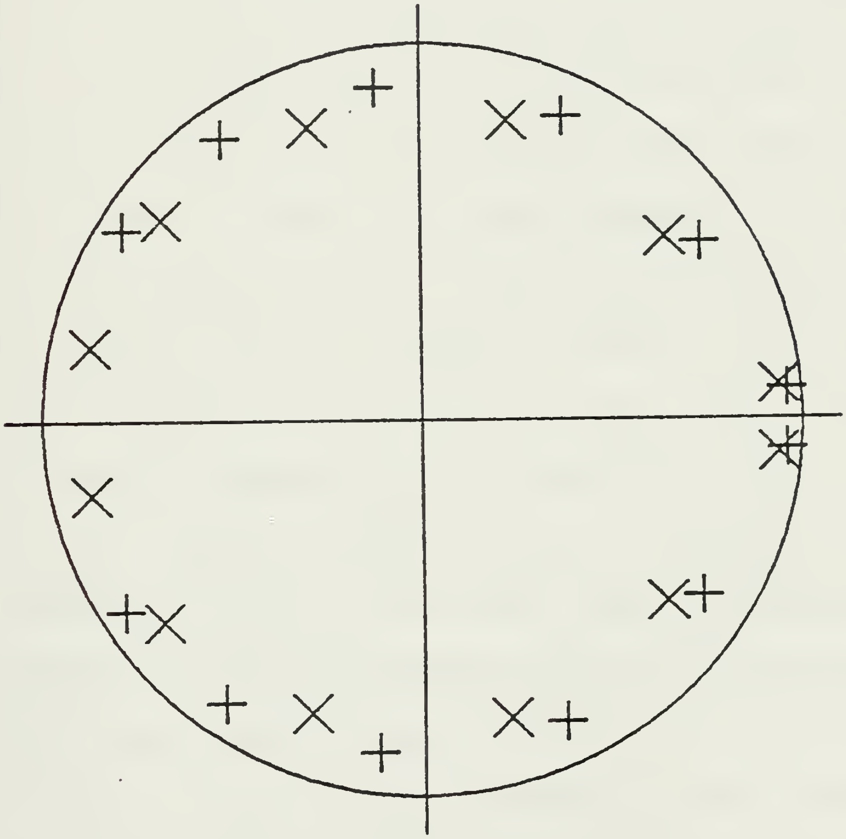


FIGURE 20. VOCAL TRACT POLES
X INPUT
+ AFTER MODIFICATION

Subroutine NEWCF performed the task of multiplying the poles to calculate the coefficients of the modified

characteristic equation for the vocal tract filter. This operation was done in double precision arithmetic because the predictor coefficients being calculated often differed by only small amounts. This process would require close study before this system could be implemented on a short word length processor.

2. Pitch Period Modification

The pitch period was modified in the main program and consisted only of converting the pitch period (an integer) to floating point representation, multiplying by the pitch period scale factor, and reconverting to fixed point representation. Although changing the pitch period is relatively simple, a number of other changes are caused by modifying the pitch period. If the pitch period is shortened the gain must be reduced to make up for the increased energy being input to the vocal tract filter. The relationship between the pitch period and the formant bandwidth also requires further study. It appears that the formant bandwidths (Q's of the vocal tract resonators) should produce a impulse response which is significantly attenuated by the time the next impulse is input to the filter. There is most likely a feedback effect between the vocal tract resonators and the vocal cords vibration rate which is not considered by the model used. This effect is noted in the graphical output as sharp discontinuities at the point where each new impulse is generated.

3. Gain Adjustment

Although overall gain of the system can be adjusted easily at the output, the relative amplitude from frame to frame must be retained during the processing. The gain coefficient, root mean square of the error function, is adjusted to account for the change in the energy of the vocal tract impulse response brought about by the bandwidth changes. As was described earlier the ratio of the original and modified vocal tract filter gain a zero frequency is used to estimate the ratio of inpulse response energy. Although this is not strictly true, as long as the scaling factors are limited to those which produce realistic speech sounds, this appears to work very well. The zero frequency gain of the original vocal tract filter, G(in), is calculated before the LPC coefficients are modified.

$$G(in) = \sum_{i=0}^{p} a_i$$

The value of both a_o and a'_o is unity. After the coefficients are modified the same calculation is performed again.

$$G(out) = \sum_{i=0}^{p} a'_i$$

The root mean square of the error signal, rms(E), is multiplied by the ratio to obtain the new gain coefficient,

rms'(E).

$$rms'(E) = rms(E) \times G(in) / G(out)$$

G. SPEECH RECONSTRUCTION

Reconstruction of the sampled speech waveform, from the modified LPC parameters is accomplished by subroutine RECCN. This routine not only decodes both voiced and unvoiced speech, but also makes allowance for the transition of varying parameters from frame to frame. The LPC parameters from the previous frame are saved between calls to subroutine COEFF and are used during the current frame when needed. It is also necessary to save output values from the previous frame to allow the recursive calculation of the output values at the begining of the current frame.

1. Unvioced Speech

During continuous unvoiced speech (as opposed to the previous frame being voiced) the new LPC parameters are used immediately upon entry to subroutine RECON. The excitation function is determined by calling a library routine GGNOF which returns normally distributed random numbers with zero mean and a variance of unity, and multiplying the value returned by the gain parameter. The excitation function is changed for every output sample to simulate the continuous excitation caused by turbulent air in the vocal tract. The vocal tract filter is implemented by the recursive addition of past values of the output to

the excitation function. The z-domain transfer function

$$\frac{s(z)}{e(z)} = \frac{1}{1 + \sum_{i=1}^{p} a_i z^{-i}}$$

is implemented with the discrete time function

$$s(n) = e(n) - \sum_{i=1}^{p} a_i \, s(n-i)$$

where s(n) is the output sample and e(n) is the excitation function.

2. Voiced Speech

During voiced speech a certain amount of continuity must be maintained from frame to frame. This was accomplished by allowing any uncompleted pulses from the previous frame to finish before the parameters are changed. Immediately upon entering the subroutine during voiced speech the pulse period counter is tested to see if it is equal to the former pulse period. If the former pulse is not complete the routine goes ahead and recursively calculates the output values. Upon completion of a pulse from a former frame or any pulse during the current frame, the new LPC parameters are used to replace the old one. There was a direct replacement for all parameters except the gain coefficient. The geometric mean of the old and new gain coefficients is used for the gain on the current pulse and the old gain replaced with the gain just calculated. This provides for the difference between the old and new

gain parameters to decay exponentially but prevents sharp changes in amplitude from frame to frame and make the output speech more natural.

3. Transition Frames

If the current frame and the previous frame were not of the same type care must be taken to insure that all parameters are changed together. If LPC coefficients for unvoiced speech were used with a pulsed output an unnatural sound would be likely to be produced. During the transition from unvoiced to voiced speech, the retained values from the previous frame are normally small in comparison to the amplitude of the pulsed excitation function. Therefore the voiced speech production may begin immediately. When the opposite is true, the large amplitude samples near the begining of a output pulse are significantly larger than the unvoiced excitation values. Therefore whenever unvoiced speech follows a voiced frame, the previous output pulse is allowed to finish. The damping that occurs during the voiced pulse normally reduces the magnitude of the samples near the end of the pulse to the point where they will not interfere with the unvoiced speech to follow.

H. OUTPUT PROCESSING

The reconstruced speech samples are output onto a standard nine track IBM 360 magnetic tape. These values were later input to a data conversion program (Appendix

A.4) which converted the floating point values to integers which were in the proper format for the XDS 9300 and within an appropriate range for the XDS 9300's digital to analog converter. The necessity of using a seven track tape for data transfer still existed, so the significant bit of the integers had to be shifted into the proper pcsition so that none of the eight bits dropped during the writing of each value onto the seven track tape would effect the data. This tape was input to the XDS 9300 which via the digital to analog converter made the samples available on the COMCOR 5000 in analog form.

These samples were output at a rate of 5000 per second thru a sample and hold circuit. Again two low pass filters were used to remove the time quantization noise from the samples. The analog waveform was recorded at 3 3/4 ips on a standard tape recorder which could be played at 7 1/2 ips to hear the reconstruced speech.

I. GRAPHICAL OUTPUT

The programs described above were also able to produce a varity of graphical outputs to assist the researcher in following the signals through the LPC processing. The waveforms available from these programs are:

(1) Input speech

(2) Error signal before filtering

(3) Error signal after filtering

(4) Reconstructed output speech

The z-plane pole locations determine the formant frequencies and bandwidths and were also available for graphical display. A seperate program (Appendix A.3) was written to display the logarithmic power spectral density of the input and output speech for a number of consecutive frames and proved useful in analysis of the output quality.

VII. RESULTS

The desired result of this study was the reconstruction of speech at different pitch and formant frequencies than that of the input speech. The complete process of encoding, modification and decoding was accomplished for three 5-second segments of speech. Upon completion of the process most listeners agreed that although the input speech was female, the modified output speech sounded typically male. Although the audio output was somewhat lacking in quality it was intelligible.

Examples of the printed and graphical computer output are given in Appendix B. Two examples are completely covered. The first 384 msec. segment (15 frames) is of the vowel 'e' and the second segment is of the transition from a fricative to a voiced sound, 'sa', from the begining of the word salt. Both were derived from a recording of a female speaker were reconstructed first without modification and then with modificaations which consisted of reduction of the pitch frequency by a factor of 0.58 and reduction of the formant frequencies by a factor of 0.88. First the input waveform with the logarithmic power spectral density plot of that portion of the speech is given. Examples of the printed processing summary are next and are followed by the waveforms of the error signal and the filtered error signal. Plots of the vocal tract pole

locations are shown with the poles at input superimposed on the poles after modification. Finally, speech waveforms for both unmodified and modified output with their respective logarithmic power spectral density functions are displayed. The audio output is available from the author on request, in the form of an audio tape recording. This tape recording is described in detail in Appendix C.

The results above demonstrate the feasibility of the use of linear predictive coding as a technique for voice modification. This research also indicated areas in which further study and improvement may be made. Some of these areas are:

(1) The effect of noise during voiced speech on the prediction error and on the gain calculated from the error. It may be possible to use only the energy occuring at the peaks of the error signal and thereby attribute the remainder of the error signal as being due to noise.

(2) The effect of the use of different window functions in autocorrelation function calculation and how this variation effects pitch period determination and the voicing threshold.

(3) The possibility of constructing a LPC processing system with asyncronous clocks for the frame timer and the output sample gereration. This would produce a very similar effect to that accomplished here, but probably at a reduced cost.

VIII. CONCLUSIONS

With the refinement and standardization of LPC commuication processors, the ratio of processing time to real time for unaltered communication is expected to drop below the current 65%. The available computation time may be used for the pitch and formant alteration described above or for other modification which can be accomplished at either the transmitting or receiving processor and still allow real time voice communications.

A number of possible applications of the speech frequency characteristic modification described are:

(1) A digital hearing aid for persons (such as the author) with high frequency hearing loss.

(2) Radios in military vehicles which would produce speech in a frequency range different than the range of the predominant noise in the vehicle, i.e. low pitch voice in turbine aircraft with high frequency noise and high pitched voice for helicopters and tanks where low frequency noise is most prevalent.

(3) Voice channel jammers which would produce random phonemes with pitch and formant characteristics similar to the current users of the channel.

As LPC communications systems become common because of their low data rate requirements, the use of the LPC parameter modification will be desired to extend the flexiblity of voice communication and storage systems. Frequency modification is one viable process available.

APPENDIX A.1 SEVEN TRACK TO NINE TRACK TAPE CONVERSION PROGRAM

```
      DIMENSION  IDAT(1024),DAT(53248)
      FACTOR = 100.0/(2.0**23)
      REWIND 2
      REWIND 4
      N=1024
      K=0
5     K=K+1
      IF(K.GT.13) STOP
6     BSUM = 0.0
7     J=0
8     J=J+1
      IF(J.LE.50) GO TO 10
12    READ(2,15,END=200) IDAT
      GO TO 12
10    READ(2,15,END=200,ERR=60) IDAT
15    FORMAT(128(8A4))
17    CALL  FORM(IDAT,N)
      JJ=(J-1)*1024
      SUM = 0.0
      DO 20 I=1,1024
      II = I+JJ
      DAT(II) = FLOAT(IDAT(I))*FACTOR
      SUM = SUM + DAT(II)
20    CONTINUE
      SUM = SUM/1024.
      WRITE(6,25) J,K
25    FORMAT(40X,'* RECORD ',I3,' OF FILE ',I3,
     *   ' HAS BEEN READ *')
      IF (J.LE.1) WRITE(6,30) K,SUM,(DAT(L),L=1,1024)
30    FORMAT(' FILE =',I3,' AVG = ',E16.7//(1X,8E14.7))
      IF (J.LE.1) WRITE(6,31) IDAT
31    FORMAT(1X,8I15)
      BSUM = BSUM + SUM
90    GO TO 8
200   J=J-1
      WRITE(6,205) K,J
205   FORMAT(' END OF FILE ',I3,I6,
     *   ' RECORDS HAVE BEEN READ')
      BSUM = BSUM/FLOAT(J)
      DO 95 J=1,51200
      DAT(J) = DAT(J)-BSUM
95    CONTINUE
      WRITE(4,98) (DAT(L),L=1,51200)
98    FORMAT(128A4)
      ENDFILE 4
      WRITE(6,30) K,BSUM,(DAT(L),L=1,1024)
100   GO TO 5
60    WRITE(6,65) K
65    FORMAT(' ** ERR FILE',I3,' **')
      GO TO 17
      END
```

APPENDIX A.2   LINEAR PREDICTIVE CODING AND VOICE
MODIFICATION PROGRAM

```
C
C   LINEAR PREDICTIVE CODING AND SPEECH MODIFICATION
C   PROGRAM
C
C   SAMPLED SPEECH IS INPUT VIA FILE FT02F001 (TAPE OR
C   DISK) IN FORMAT 128A4 FOR EFFICIENT STORAGE
C
C   SPEECH IS ENCODED INTO LPC CONSISTING OF PITCH PERIOD
C   (IPP), VOICED/UNVOICED DECISION (IVF), GAIN FACTOR
C   (RMSE), AND LPC COEFFICIENTS (A(I))
C
C   MODIFICATIONS TO CHANGE POLE POSITIONS MAY BE SPECIFIED
C   SAMPLED SPEECH IS RECONSTRUCTED AND OUTPUT ONTO FILE
C   FT03F001, ALSO IN 128A4 FORMAT
C
C   PROGRAMMED BY G.T.HALL, 1978
C
      DIMENSION X(256),A(14),XX(14),E(256),XO(256)
      DIMENSION EF(256),ES(5),EFS(5),ZERO(256)
      COMPLEX P(14)
      DATA XX,EFS,ES,ZERO/280*0.0/
C
C   SET VOICE/UNVOICE THRESHOLD
C
      THRESH = 2.05
      IWIN = 1
C
C   SET ORDER OF PREDICTOR
C
      IP = 12
C
C   PLOTTER OUTPUT
C     1=INPUT              2=ERROR SIGNAL
C     3=FILTERED ERROR     4=OUTPUR
C     5=POLE LOCATION (FIRST NPLPLT FRAMES)
C
      IXPLT = 5
      NPLPLT = 10
C
C   SET IWRXX=1 FOR PRINTED RESULTS FROM SUB
C
      IWR = 1
      IWRERR = 0
      IWRAUT = 0
      IWRALT = 1
      IWRPP = 0
      IWRPOL = 0
      IWRNC = 1
C
C   SET MODIFICATIONS DESIRED
C     (FSC) FREQUENCY SCALE COEFF
C     (BSC) BANDWIDTH SCALE COEFF
C     (PSC) PITCH PERIOD SCALE COEFF
C     (RSC) REAL POLE SCALE COEFF
C     (RLIM) REAL POLE MAGNITUDE LIMIT
C     (SP) SAMPLING INTERVAL
C
      FSC = 0.88
      BSC = 0.63
      PSC = 1.73
      RSC = 1.0
      RLIM = 0.95
      SP = 0.0001
C
C   SET NUMBER OF SAMPLES PER FRAME
C
      N = 256
```

```
C
C     SET NUMBER OF FRAMES (NFRAME) AND NUMBER OF
C     FRAMES SKIPPED BEFORE FIRST ANALYZED
C
      NFRAME = 15
      ISKIP = 28
      IF (ISKIP.LE.0) GO TO 2
      DO 1 L = 1,ISKIP
      READ (2,15,END = 999) (X(J),J=1,N)
1     CONTINUE
2     IF(IXPLT.LT.5.AND.IXPLT.GT.0) CALL VPLTIN(N)
      DO 200 I = 1,NFRAME
      READ (2,15,END = 999) (X(J),J=1,N)
15    FORMAT(128A4)
      IF (IXPLT.EQ.1) CALL VPLT(X)
C
C     DETERMINE RMS VALUE OF SPEECH SAMPLES
C
      CALL RMS (X,N,RMSX)
      IF (IWR.EQ.1) WRITE (6,20) I,RMSX
20    FORMAT('1FRAME ',I4//1X,'RMS VALUE OF SAMPLES  = ',
     * F18.8)
C
C     DETERMINE PREDICTOR COEFF BY AUTOCORRELATION METHOD
C
      CALL AUTO (X,N,A,IP,IWIN,IWRAUT)
      IF(IWR.EQ.1) WRITE(6,21) ((J,A(J)),J=1,IP)
21    FORMAT(/1X,'PREDICTOR COEFFICIENTS'/(10X,I3,1X,F18.8))
C
C     DETERMINE ZERO FREQ GAIN OF VOCAL TRACT TRANS FCN
C
      GIN = 1.0
      DO 22 J = 1,IP
      GIN = GIN + A(J)
22    CONTINUE
      GIN = 1.0/GIN
      IF (IWR.EQ.1) WRITE(6,23) GIN
23    FORMAT(/' G IN =',F10.5)
C
C     DETERMINE POLES OF CHARACTERISTIC EQUATION
C
      CALL POLES (A,IP,P,IWRPOL,ICK)
C
C     INVERSE FILTER SAMPLES TO GET ERROR SIGNAL
C
      CALL ERR (X,N,A,IP,E,XX)
      IF (IXPLT.EQ.2) CALL VPLT(E)
      IF (IWRERR.EQ.1) WRITE (6,25) (E(J),J = 1,N)
25    FORMAT(1X,10F12.4)
C
C     DETERMINE RMS VALUE OF ERROR
C
      CALL RMS (E,N,RMSE)
      IF (IWR.EQ.1) WRITE (6,30) RMSE
30    FORMAT(/1X,'RMS VALUE OF ERROR  = ',F18.8)
      RATIO  = RMSX/RMSE
      IF (IWR.EQ.1) WRITE (6,40) RATIO
40    FORMAT(/1X,'RATIO SAMPLE RMS TO ERROR RMS = ',F18.8)
C
C     TEST  IF VOICED OR UNVOICED
C
      IVF = 0
      IF (RATIO.GE.THRESH) IVF = 1
      IF (IVF.EQ.1) WRITE (6,41)
41    FORMAT(/' THIS FRAME IS VOICED'/)
      IF (IVF.EQ.0) WRITE (6,42)
42    FORMAT(/' THIS FRAME IS UNVOICED'/)
C
C     IF UNVOICED BYPASS PITCH DETECTION
C
      IF (IVF.EQ.0) GO TO 45
      CALL PITCH (N,E,EF,ES,EFS,IPP,IWRPP)
```

```
      IF (IXPLT.EQ.3) CALL VPLT(EF)
      GO TO 49
C
C  IF UNVOICED ZERO SAVED POST FILTER ERRCR
C
45    DO 46 J = 1,5
      EFS(J) = 0.0
46    CONTINUE
      IF (IXPLT.EQ.3) CALL VPLT(ZERO)
C
C  DETERMINE NEW PITCH PERIOD
C
49    IPPN = IFIX(FLOAT(IPP)*PSC+0.5)
C
C  ALTER POLE LOCATIONS
C
      IF(I.EQ.1.AND.IXPLT.EQ.5) CALL PLOTS(IA,IB,IC)
      IF(I.EQ.NPLPLT.AND.IXPLT.EQ.5) IXPLT=0
      CALL ALT2(P,FSC,BSC,RSC,RLIM,SP,IP,IWRALT,IXPLT)
      WRITE(6,51) IPPN
51    FORMAT(/' PITCH PERIOD AFTER MODIFICATION',I3)
C
C  CALCULATE NEW PREDICTOR COEFFICIENTS
C
      CALL NEWCF(IP,P,A,IWRNC)
      DO 50 J = 1,IP
      JJ = J+N-IP
      XX(J) = X(JJ)
50    CONTINUE
C
C  DETERMINE ZERO FREQ GAIN OF VOCAL TRACT TRANS FCN
C
      GOUT = 1.0
      DO 52 J = 1,IP
      GOUT = GOUT+A(J)
52    CONTINUE
      GOUT = 1.0/GOUT
      IF (IWR.EQ.1) WRITE(6,53) GOUT
53    FORMAT(/' G OUT =',F10.5)
C
C  ADJUST OUTPUT GAIN
C
      RMSE = RMSE*GIN/GOUT
      CALL RECON(A,IP,RMSE,IVF,IPPN,N,XO)
      IF (IWR.EQ.1) WRITE (6,54) (XO(L),L = 1,N)
54    FORMAT(/' OUTPUT SAMPLES'/(1X,10F13.5))
      IF (IXPLT.EQ.4) CALL VPLT(XO)
      WRITE(3,15) (XO(J),J=1,N)
200   CONTINUE
999   IPEN = 999
      CALL PLOT(A,B,IPEN)
      STOP
      END
```

```
      SUBROUTINE AUTO (S,N,A,IP,IWIN,IWR)
C
C   DETERMINE LINEAR PREDICTION COEFFICIENTS FOR A SET OF
C   INPUT SAMPLES USING THE AUTOCORRELATICN METHOD
C
C   S = VECTOR OF INPUT SAMPLES
C   N = NUMBER OF SAMPLES
C   A = VECTOR OF PREDICTOR COEFFICIENTS
C   IP = NUMBER OF PREDICTCR CCEFF ( ORDER CF MODEL  )
C             IP.LT.17
C
C   IWIN = TYPE OF WINDOW ( SEE SUBR WINDOW )
C   IWR = 0 NO PRINTING OF PREDICTION COEFFICIENTS
C
C   REF: MAKHOUL: LINEAR PREDICTION
C                 PROC IEEE, APR 75
C
      DIMENSION S(1),T(512),R(16),A(1)
      CALL WINDW (S,T,N,IWIN)
C
C   CALCULATE AUTOCORRECATICN
C
      RO = 0.0
      DO 10 I=1,N
      RO = RO + T(I)**2
10    CONTINUE
      DO 30 J=1,IP
      SUM = 0.0
      NN = N-J
         DO 20 I=1,NN
         SUM = SUM+T(I)*T(I+J)
20       CONTINUE
      R(J) = SUM
30    CONTINUE
      IF(IWR.EQ.1) WRITE(6,31) RO,(R(L),L=1,IP)
31    FORMAT(/1X,'AUTOCOREL VALS',F16.5/1X,8F16.5/1X,8F16.5)
C
C   SOLVE MATRIX EQN FOR A VECTOR
C
      CALL COEFF(RO,R,IP,A,IWR)
C
C   TAKE NEGITIVE OF PREDICTOR COEFF TO GET
C   COEFF OF CHARACTERISTIC EQN OF FILTER
C
      DO 60 I = 1,IP
      A(I) = -A(I)
60    CONTINUE
      IF (IWR.NE.0) WRITE(6,70) ((I,A(I)),I=1,IP)
70    FORMAT(/1X,'PREDICTOR CCEFFICIENTS'/(10X,I3,1X,F18.8))
      RETURN
      END
```

```
      SUBROUTINE COEFF(RO,R,N,A,IWR)
C
C   SOLVES THE MATRIX EQUATION RR A = R
C
C   RR =   AUTOCORRELATION MATRIX
C   RR =   R(0)    R(1)    R(2).......R(N-1)
C          R(1)    R(0)    R(1).......R(N-2)
C          R(2)    R(1)    R(0).......R(N-3)
C            .       .       .              .
C          R(N-1)  R(N-2)  R(N-3).....R(0)
C
C   R =     AUTOCORRELATION VECTOR
C   R =    R(1)
C          R(2)
C          R(3)
C            .
C          R(N)
C
C   A = VECTOR OF PREDICTOR COEFF
C   A = A(1)
C       A(2)
C       A(3)
C         .
C       A(N)
C
C   METHOD ATTRIBUTED TO DURBIN AS DESCRIBED IN
C   'LINEAR PREDICTION' BY MAKHOUL,PROC IEEE APR 75
C   P. 566
C
      DIMENSION AK(20),AO(20),A(20),R(20)
C
C     FIRST ITERATION
C
      EO = RO
      AK(1) = R(1)/EO
      A(1) = AK(1)
      E = (1.0-AK(1)**2)*EO
      EO = E
      AO(1) = A(1)
C
C     FOLLOWING ITERATIONS
C
      DO 100 I = 2,N
      IM1 = I-1
      SUM = 0.0
         DO 20 J = 1,IM1
         IMJ = I-J
         SUM = SUM+R(IMJ)*AO(J)
20       CONTINUE
      AK(I) = (R(I)-SUM)/EO
      A(I) = AK(I)
         DO 30 J = 1,IM1
         IMJ = I-J
         A(J) = AO(J)-AK(I)*AO(IMJ)
30       CONTINUE
      E = (1.0-AK(I)**2)*EO
      EO = E
      DO 50 J = 1,I
      AO(J) = A(J)
50    CONTINUE
100   CONTINUE
C
C   PRINT E (REMAINING ERROR DUE TO LIMITING
C   ORDER OF APPROXIMATION) AND A CHECK OF SOLUTION
C   IF DESIRED
C
      IF(IWR.EQ.1) WRITE(6,101) E
101   FORMAT(' SUB COEFF E= ',F18.8)
      IF(IWR.EQ.1) CALL TEST(A,RO,R,N)
      RETURN
      END
```

```
      SUBROUTINE TEST (A,RO,R,IP)
C
C   MULTIPLIES PREDICTOR CCEFF VECTOR
C   A BY THE AUTOCORRELATION MATRIX RR AND CHECKS
C   THE VALUE AGAINST THE AUTOCORRELATION VECTOR
C   TO INSURE ACCURATE SOLUTION.
C
      DIMENSION A(IP),R(IP)
      DO 1C I = 1,IP
      SUM = 0.0
         DO 9 J = 1,IP
         L = IABS(I-J)
         IF(L.EQ.0) SUM = SUM+A(J)*RO
         IF(L.NE.0) SUM = SUM+A(J)*R(L)
9        CONTINUE
      WRITE(6,15) I,R(I),SUM
15    FORMAT(' R(',I2,') = ',2E14.4,' = SUM')
10    CONTINUE
      RETURN
      END
```

```
      SUBROUTINE POLES (A,IP,P,IWR,ICK)
C
C   CALCULATES POLES OF CHARACTERISTIC EQN FROM
C   PREDICTOR COEFFICIENTS AND IF WANTED PRINTS
C   OR PLOTS THOSE POLES
C
C   A = VECTOR OF PREDICTOR COEFFICIENTS
C   IP = NUMBER OF CCEFF AND POLES
C            COEFF A0 IS ASSUMED TO BE 1.0
C   P = COMPLEX VECTOR OF POLE LOCATIONS
C   IWR = 0 NO PRINTING OF POLES
C   ICK = 0 ALL POLES INSIDE UNIT CIRCLE
C       = 1 POLE OUTSIDE UNIT CIRCLE
C
      DIMENSION A(1),B(21),X(20),Y(20),NAME(20)
      COMPLEX P(1)
      B(1) = 1.0
      DO 10 I=1,IP
      II = I+1
      B(II) = A(I)
10    CONTINUE
      IIP = IP+1
      CALL ZRPOLY(B,IP,P,IER)
      IF(IWR.NE.0) WRITE(6,20) ((I,P(I)),I=1,IP)
20    FORMAT(//10X,'POLES OF CHAR EQN'/(10X,I3,1X,2E14.7))
      ICK = 0
      DO 30 I = 1,IP
      IF(CABS(P(I)).LE.1.0) GO TO 30
      ICK = 1
      IF(IWR.NE.0) WRITE(6,25) I
25    FORMAT(20X,'POLE NUMBER ',I3,
     *  ' ABOVE IS OUTSIDE UNIT CIRCLE')
30    CONTINUE
      RETURN
      END
```

```
      SUBROUTINE ERR (S,N,A,IP,E,SX)
C
C   DETERMINE AN ERRCR VECTOR CF DIFFERENCE
C   BETWEEN ACTUAL SAMPLE VALUES AND THE
C   VALUES PREDICTED FROM PAST SAMPLES.
C
C   S = VECTOR OF SAMPLES
C   N = NUMBER OF SAMPLES
C   A = VECTOR OF PREDICTOR COEFF
C   IP = NUMBER OF PREDICTOR COEFF
C   E = VECTOR OF ERROR VALUES
C   SX =  EXTRA SAMPLES ( IP OF THEM )
C         SAVED FROM LAST FRAME
C
C   THE ERROR IN THE DIFFERENCE BETWEEN THE
C   CURRENT SAMPLE AND THE WEIGHTED SUM OF
C   THE LAST IP SAMPLES.
C
      DIMENSION S(1),A(1),E(1),T(542),SX(1)
      DO 10 I=1,IP
      T(I) = SX(I)
10    CONTINUE
      DO 20 I=1,N
      T(I+IP) = S(I)
20    CONTINUE
      DO 40 I=1,N
      SUM = 0.0
         DO 30 J=1,IP
         II = I+J-1
         JJ = IP-J+1
         SUM = SUM+T(II)*A(JJ)
30       CONTINUE
      E(I) = S(I) + SUM
40    CONTINUE
      RETURN
      END
```

```
      SUBROUTINE PITCH(N,E,EF,ES,EFS,IPP,IWR)
C
C   DETERMINES PITCH PERIOD (IN NUMBER OF SAMPLES)
C   FROM THE ERROR SIGNAL OF INVERSE FILTERED SPEECH
C
C   N = NUMBER OF SAMPLES
C
C   E = ERROR VECTOR
C
C   EF = FILTERED ERROR VECTOR (OUTPUT)
C
C   ES = FIVE SAVED ERROR SAMPLES
C
C   EFS = FIVE SAVED FILTERED ERROR SAMPLES
C
C   IPP = PITCH PERIOD (OUTPUT)
C
C   IWR = 1 FOR PRINTING DURING SUBROUTINE
C
C
      DIMENSION ES(5),EFS(5),E(1),EF(1),R(256)
      DIMENSION XI(261),XO(261)
C
C    FORM FILTERING VECTOR(N+5)
C
      DO 10 I=1,5
      XI(I)=ES(I)
      XO(I)=EFS(I)
10    CONTINUE
      ITEMP=N+5
      DO 15 I=6,ITEMP
      II=I-5
      XI(I)=E(II)
15    CONTINUE
      DO 20 I=6,ITEMP
C
C     BUTTERWORTH DIGITAL FILTER CUTOFF AT 800 HZ
C
      XO(I) = 0.447451239E-3*XI(I)+C.22372562E-2*XI(I-1)
     *         +0.44745124E-2*XI(I-2)+0.447451239E-2*XI(I-3)
     *         +0.22372562E-2*XI(I-4)+0.447451239E-3*XI(I-5)
     *         +3.41077231*XO(I-1)-4.7828C887*XO(I-2)
     *         +3.42533523*XO(I-3)-1.24929545*XO(I-4)
     *         +0.185257941*XO(I-5)
C
      EF(I-5)=XO(I)
20    CONTINUE
      DO 30 I=1,5
      ES(I)=E(I+N-5)
      EFS(I)=EF(I+N-5)
30    CONTINUE
      IWIN=4
      CALL WINDW(EF,XO,N,IWIN)
C
C   CHECK FOR PEAKS 1.2 TO 18.C MSEC
C
      ITEMPC=N-56
      IF(IWR.EQ.1) WRITE(6,33) ((EF(L),XO(L)),L=1,N)
33    FORMAT(1X,10F13.5)
      DO 50 I=1,ITEMPC
      SUM=0.0
      ITEMPA=N-I
         DO 40 J=1,ITEMPA
         SUM=SUM+XO(J)*XO(J+I)
40       CONTINUE
      R(I)=SUM
      IF(IWR.EQ.1) WRITE(6,41) I,R(I)
41    FORMAT(' FILTERED ERROR AUTOCORRELATION FOR',I4,
     *   F18.8)
      IF(I.LT.33) GC TO 50
      ITEST=I-25
      IF(R(ITEST).LT.0.0) GO TO 50
```

```
      ITEMPB=I-34
         DO 45 J=ITEMPB,I
         IF(R(ITEST).LT.R(J)) GO TO 50
45       CONTINUE
      IPP=ITEST
      WRITE(6,46) IPP
46    FORMAT (' PITCH PERIOD IS',I4)
      RETURN
50    CONTINUE
      IPP=100
      WRITE(6,55)
55    FORMAT(///' SUB PITCH FAILED TO DETERMINE CORRECT'
     * /' PITCH, PITCH PERIOD SET EQUAL TO 100'///)
      RETURN
      END
```

```
      SUBROUTINE ALT2 (P,FSC,BSC,RSC,RLIM,SP,IP,IWR,IXPLT)
C
C   GIVEN IP COMPLEX POLES OF THE VOCAL TRACT
C   TRANSFER FUNCTION, CALCULATES THE FORMANT
C   FREQUENCIES AND BANDWIDTHS AND SCALES THEM
C   AS DESIRED. PRINTED OUTPUT IS AVAILABLE.
C
C   P = VECTOR OF IP COMPLEX POLES
C   FSC = FREQUENCY SCALE FACTOR OUT/IN
C   RSC = REAL POLE SCALE FACTOR
C   RLIM = REAL POLE MAGNITUDE LIMIT
C   BSC = BANDWIDTH SCALE FACTOR OUT/IN
C   IP = NUMBER OF POLES
C   SP = SAMPLE PERIOD IN SECONDS
C   IWR = 0 NO OUTPUT PRINTED
C         1 PRINTED RESULTS
C   IXPLT = 5 FOR PLOT OF POLES
C
      DIMENSION FORF(14),BW(14)
      COMPLEX P(1),CPP(14),CRP(14),CTEM
      DIMENSION XP(6),YP(6),IIPEN(6)
      DATA XP/3.0,2.75,-2.75,0.0,0.0,2.5/
      DATA YP/10.0,0.0,0.0,2.75,-2.75,0.0/
      DATA IIPEN/-3,3,2,3,2,3/
      ZERO=0.0
      IF(IXPLT.NE.5) GO TO 9
C
      NPEN = 3
      CALL NEWPEN(NPEN)
      DO 2 I=1,6
      CALL PLOT(XP(I),YP(I),IIPEN(I))
2     CONTINUE
      IPEN = 2
C
      DO 4 I=1,241
      TEM = 0.02618*FLOAT(I)
      XX = 2.5 * COS(TEM)
      YY = 2.5 * SIN(TEM)
      CALL PLOT(XX,YY,IPEN)
4     CONTINUE
C
      IPEN = 3
      CALL PLOT (ZERO,ZERO,IPEN)
C
      HIEG = 0.25
      ANG = 0.0
      NC = -1
      ITEXT = 4
      NPEN = 4
      CALL NEWPEN(NPEN)
C
      DO 6 I=1,IP
      XX = 2.5 * REAL(P(I))
      YY = 2.5 * AIMAG(P(I))
      CALL SYMBOL(XX,YY,HIEG,ITEXT,ANG,NC)
6     CONTINUE
C
9     IRP = 0
      ICP = 0
C
C   TEST EACH POLE AND PLACE IN PROPER ARRAY
C
      DO 40 I=1,IP
      IF(AIMAG(P(I)).EQ.0.0) GO TO 30
      IF(ICP.EQ.0) GO TO 20
         DO 10 J=1,ICP
         IF(CABS(P(I)-CONJG(CPP(J))).LT.0.001) GO TO 40
10       CONTINUE
20    ICP=ICP+1
      CPP(ICP)=P(I)
      GO TO 40
30    IRP=IRP+1
```

```
      CRP(IRP)=P(I)
40    CONTINUE
C
C  CALCULATE FORMANT FREQ AND BANDWIDTH FOR EACH
C
      DO 50 I=1,ICP
      A=CABS(CPP(I))
      BW(I)=(0.0-ALOG(A))/(6.2831852*SP)
      TH=ATAN2(AIMAG(CPP(I)),REAL(CPP(I)))
      TH=ABS(TH)
      FORF(I)=TH/(SP*6.2831852)
50    CONTINUE
      ICPM1=ICP-1
      DO 60 I=1,ICPM1
      IP1=I+1
         DO 55 J=IP1,ICP
         IF(FORF(I).LT.FORF(J)) GO TO 55
         TEM=BW(I)
         BW(I)=BW(J)
         BW(J)=TEM
         TEM=FORF(I)
         FORF(I)=FORF(J)
         FORF(J)=TEM
         CTEM=CPP(I)
         CPP(I)=CPP(J)
         CPP(J)=CTEM
55       CONTINUE
60    CONTINUE
      IF(IWR.EQ.1) WRITE(6,70) ((I,CPP(I),FORF(I),BW(I)),
     * I=1,ICP)
70    FORMAT(' FORMANT',I3,' DUE TO POLES AT Z=',F8.4,
     * '+-J*',F8.4,' FORMANT FREQ=',F8.1,' BANDWIDTH=',F8.1)
      IF (IRP.EQ.0) GO TO 85
      IF(IWR.EQ.1) WRITE(6,80) ((I,CRP(I)),I=1,IRP)
80    FORMAT(' REAL POLE NUMBER ',I3,' AT Z =',2F8.4)
85    IF(IWR.EQ.1) WRITE(6,90) FSC,BSC,RSC,RLIM,SP
90    FORMAT(/' FORMANT FREQUENCY SCALE FACTOR =',F8.4,
     *    '     BANDWIDTH SCALE FACTOR =',F8.4/
     *    ' REAL POLE SCALE FACTOR =',F8.4,
     *    '   REAL POLE MAGNITUDE LIMIT =',F8.4,
     *    '   SAMPLE PERIOD =',F9.6//' AFTER MODIFICATION')
C
C  ALTER FORMANT FREQUENCIES AND BANDWIDTHS
C
      DO 100 I=1,ICP
      A=CABS(CPP(I))**BSC
      IF(A.GT.0.98) A=0.99
      TH=ATAN2(AIMAG(CPP(I)),REAL(CPP(I)))*FSC
      TH=ABS(TH)
      CPP(I)=A*CMPLX(COS(TH),SIN(TH))
      BW(I)=(0.0-ALOG(A))/(6.2831852*SP)
      FORF(I)=TH/(6.2831852*SP)
100   CONTINUE
C
C  ALTER REAL POLE LOCATIONS
C
      IF (IRP.EQ.0) GO TO 115
      DO 110 I=1,IRP
      CRP(I)=CRP(I)*RSC
      TEM=CABS(CRP(I))
      IF(TEM.GT.RLIM) CRP(I)=CRP(I)*RLIM/TEM
110   CONTINUE
115   IF(IWR.EQ.1) WRITE(6,70) ((I,CPP(I),FORF(I),BW(I)),
     *  I=1,ICP)
      IF (IRP.EQ.0) GO TO 118
      IF(IWR.EQ.1) WRITE(6,80) ((I,CRP(I)),I=1,IRP)
C
C  RECONSTRUCT ARRAY OF POLES
C
118   IND=0
      DO 120 I=1,ICP
      IND=IND+1
```

```
      P(IND)=CPP(I)
      IND=IND+1
      P(IND)=CONJG(CPP(I))
120   CONTINUE
      IF (IRP.EQ.0) GO TO 135
      DO 130 I=1,IRP
      INC=IND+1
      P(IND)=CRP(I)
130   CONTINUE
135   IF(IWR.EQ.1) WRITE(6,140) IND
140   FORMAT(10X,'RECON POLES',I4)
      IF(IXPLT.NE.5) RETURN
C
      ITEXT = 3
      DO 150 I=1,IP
      XX = 2.5 * REAL(P(I))
      YY = 2.5 * AIMAG(P(I))
      CALL SYMBOL(XX,YY,HEIG,ITEXT,ANG,NC)
150   CONTINUE
      IPEN = -3
      XX = 5.0
      YY = -10.0
      CALL PLOT(XX,YY,IPEN)
      RETURN
      END
```

```
      SUBROUTINE NEWCF (IP,P,A,IWR)
C
C   DETERMINES THE COEFFICIENTS OF THE
C   PREDICTOR POLYNOMIAL FROM THE ROOT OF
C   THE CHARACTERISTIC EQUATION
C
C   IP = ORDER OF THE POLYNOMIAL
C
C   P = COMPLEX ROOTS OF CHARACTERISTIC EQN
C       I.E. POLES OF THE FILTER
C
C   A = ARRAY OF REAL COEFFICIENTS
C
C   IWR = 1 FOR PRINTING DURING SUBROUTINE
C
C   IF ALL COMPLEX ROOTS ARE IN CONJUGATE PAIRS
C   ALL OF THE COEFFICIENTS SHOULD BE REAL
C   THIS CAN BE CHECKED WITH OUTPUT
C
      COMPLEX*16 PP(14),AA(14)
      COMPLEX*8 P(IP)
      REAL*4 A(IP)
      Z = 0.0
      DO 10 I = 1,IP
      AA(I) = P(I)
      PP(I) = P(I)
10    CONTINUE
      K = IP
      M = IP-1
      DO 40 L = 1,M
         DO 30 I = 2,K
         AA(I) = AA(I)+AA(I-1)
30       CONTINUE
      K = K-1
         DO 20 I = 1,K
         J = I+L
         AA(I) = PP(J)*AA(I)
20       CONTINUE
40    CONTINUE
      K = IP/2
      K = 2*K/(IP-K)
      DO 50 I = K,IP,2
      AA(I) = -AA(I)
50    CONTINUE
      DO 60 I = 1,IP
      J = IP+1-I
      A(J) =REAL(AA(I))
      PP(J) = AA(I)
60    CONTINUE
      IF(IWR.NE.1) RETURN
      WRITE(6,70) ((I,PP(I)),I = 1,IP)
70    FORMAT(/' RECONSTRUCTED POLYNOMIAL COEFFICIENTS'/
     *       20X,'IMAGINARY TERMS SHOULD BE ZERO'/
     *       (1X,I5,F18.8,E18.4))
      RETURN
      END
```

```
      SUBROUTINE RECON(A,IP,RMS,IVF,IPP,N,S)
C
C  RECONSTRUCTS SPEECH SAMPLES FROM LPC COEFF, ETC
C
C  A   =  VECTOR OF LPC COEFF
C  IP   =  NUMBER OF COEFF (ORDER OF FILTER)
C  RMS   =  RMS VALUE OF ERROR SIGNAL
C  IVF   =  0 UNVOICED
C        =  1 VOICED
C  IPP   =  PITCH PERIOD IN NUMBER OF SAMPLES
C  N   =  SAMPLES PER FRAME
C  S   =  SAMPLE VECTOR (OUTPUT)
C
      DIMENSION A(1),S(1),X(270),XX(14),AO(14)
      DATA XX,RMSO,ISEED,IVFO/15*0.0,1234,0/
      DO 10 I = 1,IP
      X(I) = XX(I)
10    CONTINUE
      NIP = N+IP
      NS  = 1+IP
C
C  IF CURRENT PULSE UNFINISHED DON'T CHANGE COEFF YET
C
      IF(IVFO.NE.0) GO TO 400
C
C  UPDATE COEFF
C
100   RMSO = SQRT(RMSO*RMS)
      IF(IVF.EQ.0) RMSO=RMS
      IF(RMSO.LT.(RMS/2.0)) RMSO=RMS/2.0
      DO 105 I = 1,IP
      AO(I) = A(I)
105   CONTINUE
      IVFO = IVF
      IPPO = IPP
C
C  TEST IF VOICED
C
      IF(IVFO.NE.0) GO TO 300
C
C  RECONSTRUCT UNVOICED SPEECH
C
200   E = RMSO*GGNOF(ISEED)
      DO 210 I = 1,IP
      NSMI = NS-I
      E = E-A(I)*X(NSMI)
210   CONTINUE
      X(NS) = E
      IF(NS.GE.NIP) GO TO 600
      NS = NS+1
      GO TO 200
C
C  START VOICED PULSE
C
300   NP = 1
      EX =RMSO*SQRT(FLOAT(IPPO))
C
C  TEST FOR BEGINING OF PULSE PERIOD
C
400   IF(NP.GT.IPPO) GO TO 100
      E = 0.0
      IF(NP.EQ.1) E = -EX
C
C  RECONSTRUCT VOICED SPEECH
C
500   DO 510 I = 1,IP
      NSMI = NS-I
      E = E-A(I)*X(NSMI)
510   CONTINUE
      NP = NP+1
      X(NS) = E
      IF(NS.GE.NIP) GO TO 600
```

```
      NS = NS+1
      GO TO 400
C
C  SAVE VALUES AND PREPARE OUTPUT
C
600   DO 610 I = 1,IP
      XX(I) = X(N+I)
610   CONTINUE
      DO 620 I = 1,N
      S(I) = X(I+IP)
620   CONTINUE
      RETURN
      END
```

```
      SUBROUTINE RMS (X,N,VAL)
C
C  DETERMINE THE RMS VALUE OF A SET OF CATA
C
C  X = VECTOR OF INPUT SAMPLES
C  N = NUMBER OF SAMPLES
C  VAL = RMS VALUE RETURNED
C
      DIMENSION X(1)
      VAL = 0.0
      DO 10 I = 1,N
      VAL = VAL+X(I)**2
10    CONTINUE
      VAL = SQRT(VAL/FLOAT(N))
      RETURN
      ENC
```

```
      SUBROUTINE WINDOW(X,Y,N,IWIN)
C
C   X = VECTOR OF UNWINDOWED SAMPLES
C   Y = VECTOR OF WINDOWED SAMPLES (OUTPUT)
C   N = NUMBER OF SAMPLES
C   IWIN = TYPE OF WINDOW
C   0 = RECTANGULAR  (COPY ONLY)
C   1 = HAMMING (ALPHA = 0.54)
C   2 = BARTLETT
C   3 = BLACKMAN
C   4 = HANNING
C
      DIMENSION X(1),Y(1)
      DATA PI,TWOPI,FORPI/3.1415926,6.2831853,12.566371/
      IF(IWIN.LT.0.OR.IWIN.GT.4) GO TO 999
      AN = FLOAT(N)
      GO TO (110,210,310,410),IWIN
C
C   RECTANGULAR WINDOW  COPY VECTOR
C
10    DO 20 I=1,N
      Y(I) = X(I)
20    CONTINUE
      RETURN
C
C   HAMMING WINDOW
C
110   DO 120 I=1,N
      AJ = FLOAT (I-1)
      Y(I) = X(I)*(0.54-0.46*COS(TWOPI*AJ/(AN-1.0)))
120   CONTINUE
      RETURN
C
C   BARTLETT WINDOW
C
210   NN = N/2
      NNN = NN+1
      DO 220 I=1,NN
      AJ = FLOAT(I-1)
      Y(I) = X(I)*2.0*AJ/(AN-1.0)
220   CONTINUE
      DO 230 I=NNN,N
      AJ = FLOAT(I-1)
      Y(I) = X(I)*2.0*(1.0-AJ/(AN-1.0))
230   CONTINUE
      RETURN
C
C   BLACKMAN WINDOW
C
310   DO 320 I=1,N
      AJ = FLOAT(I-1)
      Y(I) = X(I)*(0.42-0.5*COS(TWOPI*AJ/(AN-1.0))
     *            +0.08*COS(FORPI*AJ/(AN-1.0)))
320   CONTINUE
      RETURN
C
C   HANNING WINDOW
C
410   DO 420 I=1,N
      AJ = FLOAT(I-1)
      Y(I) = X(I)*0.5*(1.0-COS(TWOPI*AJ/(AN-1.0)))
420   CONTINUE
      RETURN
999   WRITE(6,998)
998   FORMAT(//10X,'** ERROR SUBR WINDOW **'//)
      STOP
      END
```

```
      SUBROUTINE VPLTIN (N)
C
C  SUBROUTINE CREATES A VERSAPLOT GRAPH CF 60 FRAMES
C  OF VOICE SAMPLES (128 SAMPLES / FRAME)
C
C  CALL VPLTIN TO INITIALIZE EACH PLOT
C
C  CALL VPLT FOR EACH FRAME
C
C     N=NUMBER OF SAMPLES PER FRAME
C     X=VECTOR OF SAMPLES
C
C  CALLING PROGRAM SHOULD ISSUE
C     CALL PLOT(X,Y,999)
C  TO COMPLETE PLOTTING
C
      DIMENSION X(768),Y(256),XO(8),YO(8)
      DATA XO/0.0,0.0,7.0,0.0,7.0,0.0,7.C,0.0/
      DATA YO/10.,-10.,10.,-1C.,10.,-10.,10.,-10./
      DO 10 I=1,768
      X(I)=FLOAT(I)/128.0
10    CONTINUE
      CALL PLOTS(IA,IB,IC)
      NPEN=2
      CALL NEWPEN(NPEN)
      NPLT=1
      IPEN = -3
      CALL PLOT (XO(NPLT),YO(NPLT),IPEN)
      IPEN=2
      IX=768
      IY=11
      RETURN
C
C
      ENTRY VPLT(Y)
      DO 100 I=1,N
      IX=IX+1
      IF(IX.LE.768) GO TO 50
      IX=1
      IY=IY-1
      YS=2.0+0.7*FLCAT(IY)
      IF(IY.GE.1) GO TO 4C
      NPLT=NPLT+1
      IPEN=-3
      CALL PLOT(XO(NPLT),YO(NPLT),IPEN)
      IPEN=2
      IX=1
      IY=10
      YS=2.0+0.7*FLOAT(IY)
40    IPEN=3
      YY=Y(I)/100.0+YS
      CALL PLOT(X(IX),YY,IPEN)
      IPEN=2
      GO TC 100
50    YY=Y(I)/100.0+YS
      CALL PLOT(X(IX),YY, IPEN)
100   CONTINUE
      RETURN
      END
```

APPENDIX A.3   POWER SPECTRAL DENSITY ANALYSIS AND PLOTTING PROGRAM

```
      DIMENSION X(256)
      READ(5,8,END=90) INUM,ISKIP,IWIN
8     FORMAT(3I5)
      IF(ISKIP.EQ.0) GO TO 10
      DO 9 I=1,ISKIP
      READ(2,25,END=90) X
9     CONTINUE
10    M=8
      READ(2,25,END=90) X
      CALL PSDINT(X,M)
      CALL SPLINT
      K=0
20    READ(2,25,END=90) X
25    FORMAT(128A4)
      CALL PSD(X,M,IWIN)
      IF(K.LE.6) WRITE(6,30) (X(J),J=1,128)
30    FORMAT(1X,10F12.5)
      K=K+1
      CALL SPL (X)
      IF(K.GT.INUM) GO TO 90
      GO TO 20
90    IPEN=999
      CALL PLOT(AX,Y,IPEN)
      STOP
      END
```

```
      SUBROUTINE SPLINT
C
C   SUBROUTINE PLOTS THE POWER SPECTRAL DENSITY
C   (LOG OF MAGNITUDE) FOR 128 FREQUENCIES WHICH
C   IS INPUT IN MAGNITUDE FORM IN VECTOR Y
C
C   VALUES IN Y SHOULD BE BETWEEN 0.01 AND 100.0
C
C   CALL SPLINT     TO INITIALIZE PLOTTING
C
C   CALL SPL (Y)    FOR EVERY SET OF 128 PSD VALUES
C
C   CALLING PROGRAM SHOULD ISSUE CALL PLOT (X,Y,999)
C   WHEN PLOTTING IS COMPLETE
C
C
      DIMENSION Y(1),X(128),YY(128)
      DIMENSION RORGX(6),RORGY(6),GX(19),GY(19),IGP(19)
C
C   DATA FOR SIX PLOT ORIGINS
C
      DATA RORGX/0.1,-1.2,-1.2,8.8,-1.2,-1.2/
      DATA RORGY/0.5,4.0,4.0,-17.0,4.0,4.0/
C
C   DATA TO PLOT AXIS
C
      DATA GX/7.5,7.5,6.0,6.0,4.5,4.5,3.0,3.0,1.5,
     *   1.5,0.0,0.0,-0.1,0.0,-0.1,0.0,-0.1,0.0,-0.1/
      DATA GY/0.0,-0.1,0.0,-0.1,0.0,-0.1,0.0,-0.1,
     *   0.0,-0.1,-0.1,0.800,0.800,0.600,0.600,0.4,0.4,
     *   0.200,0.200/
      DATA IGP/2,2,3,2,3,2,3,2,3,2,3,2,2,3,2,3,2,3,2/
      DO 10 I=1,128
      X(I)=FLOAT(I-1)*0.05859-0.04
10    CONTINUE
      CALL PLOTS (IA,IB,IC)
      IFLAG=0
      IPLTN=1
      ISCAN=0
15    IPEN=-3
      CALL PLOT (RORGX(IPLTN),RORGY(IPLTN),IPEN)
      NPEN=4
      CALL NEWPEN(NPEN)
      DO 30 I=1,19
      CALL PLOT(GX(I),GY(I),IGP(I))
30    CONTINUE
      NPEN=2
      CALL NEWPEN(NPEN)
      RETURN
C
C
      ENTRY SPL (Y)
      ISCAN = ISCAN + 1
C
C   RETURN IMMEDIATELY IF PLOT FULL
C
      IF(IFLAG.EQ.1) RETURN
C
C   CONVERT DATA TO LOG PLOT
C
      DO 50 I=1,128
      YTEM=Y(I)
      IF(YTEM.LT.0.100) YTEM=0.100
      YY(I)=0.10+0.2000*ALOG10(YTEM)
50    CONTINUE
      IPEN=-3
      XSCANO=0.04
      YSCANO=0.1
      CALL PLOT (XSCANO,YSCANO,IPEN)
      IPEN=3
      CALL PLOT(X(1),YY(1),IPEN)
      IPEN=2
```

```
      DO 60 I=2,128
      CALL PLOT (X(I),YY(I),IPEN)
60    CONTINUE
      IF(ISCAN.LE.29) RETURN
      ISCAN=0
      IPLTN=IPLTN+1
      IF(IPLTN.LE.6) GO TO 15
      IFLAG=1
      RETURN
      END
```

```
      SUBROUTINE PSCINT(X,M)
C
C     POWER SPECTRAL DENSITY BASED ON ALGORITHM
C     PRESENTED BY C M RADER IN, 'AN IMPRCVED
C     ALGORITHM FOR HIGH SPEED AUTOCORRELATION
C     WITH APPLICATION TO SPECTRAL ESTIMATION,'
C     IEEE TRANS AUDIO,ELECTRCACOUSTICS, V AU-18,DEC70
C
C     X = VECTOR OF INPUT SAMPLES
C     M = POWER OF 2 FOR NUMBER OF SAMPLES
C     IWIN = 0 NO WINDOW
C            1 HAMMING (ALPHA = 0.54)
C            2 BARTLETT
C            3 BLACKMAN
C            4 HANNING
C
C     FIRST CALL IS TO PSDINT AND THEN EACH SUCESSIVE
C     CALL FOR THAT STRING OF DATA SHOULC BE TO PSD
C     TO START A FRESH STRING OF DATA CALL PSDINT AGAIN
C
C
      DIMENSION X(256),IWK(11)
      COMPLEX XN(512),XNP(512),YN(512),AI(512)
      DATA XN,XNP/1C24*(C.0,C.0)/
      MM = M+1
      N=2**M
      NN=2*N
C
C     SPECIFY COEFFICIENTS NEEDED IN ADDITION
C     OF NEXT X(F) VECTOR TO CURRENT X(F) VECTOR
C     TO MAKE Y(F) VECTOR. IN BINARY REVERSE ORDER.
C
      NNN = NN-1
      DO 90 I=1,NNN,2
      AI(I) = (1.0,0.0)
      II = I+1
      AI(II) = (-1.0,0.0)
90    CONTINUE
      CALL FFRDR2 (AI,MM,IWK)
      AIMG = 0.0
      DO 101 I = 1,N
      XN(I) = CMPLX(X(I),AIMG)
101   CONTINUE
C
C     FFT OF CURRENT X(T) VECTOR,LAST HALF ZERO.
C
      CALL FFT2 (XN,MM,IWK)
      RETURN
C
C     USE THIS ENTRY FOR EACH FRAME AFTER FIRST
C
      ENTRY PSD(X,M,IWIN)
      MM = M+1
      N = 2**M
      NN = 2*N
      AN = FLOAT (N)
      ANN = FLOAT (NN)
      AIMG = 0.0
      DO 110 I = 1,N
      XNP(I) = CMPLX(X(I),AIMG)
110   CONTINUE
C
C     FFT OF NEXT X(T) VECTOR,LAST HALF ZERO.
C
      CALL FFT2 (XNP,MM,IWK)
C
C     FORM Y(F) VECTCR,COEFF IN REV BINARY ORDER.
C
      DO 120 I = 1,NN
      YN(I) = (XN(I)+AI(I)*XNP(I))*CONJG(XNP(I))
120   CONTINUE
      DO 123 I = 1,NN
```

```
C
C       FORM CONJG TO PREFORM INV DFT
C
        YN(I) = CONJG (YN(I))
123     CONTINUE
C
C       INV FFT OF Y(F) GIVES RXX(TAU)
C
        CALL FFT2RV (YN,MM,IWK)
        DO 143 I = 1,NN
        YN(I) = CONJG (YN(I))/ANN
143     CONTINUE
        CALL WIND2 (YN,N,IWIN)
        CALL FFT2 (YN,M,IWK)
        CALL FFRDR2 (YN,M,IWK)
        DO 153 I = 1,N
        X(I) = CABS (YN(I))/(AN**2)
153     CONTINUE
C
C       MOVE NEXT X(F) INTO CURRECT X(F)
C
        DO 160 I = 1,NN
        XN(I) = XNP(I)
        XNP(I) = (0.0,0.0)
160     CONTINUE
        RETURN
        END




        SUBROUTINE WIND2 (B,N,IWIN)
        COMPLEX B(512)
        DATA PI,TWOPI/3.1415926,6.283185/
        AN = FLOAT(N)
50      GO TO (200,300,400,100),I
        RETURN
100     DO 190 I = 1,N
        AJ = FLOAT(I-1)
        F = 0.5*(1.0-COS ((TWOPI*AJ)/(AN-1.0)))
        B(I) = B(I)*F
190     CONTINUE
        RETURN
200     DO 290 I = 1,N
        AJ = FLOAT(I-1)
        F = 0.54-0.46*COS ((TWOPI*AJ)/(AN-1.0))
        B(I) = B(I)*F
290     CONTINUE
        RETURN
300     DO 390 I = 1,N
        AJ = FLOAT(I-1)
        IF(I.LE.(N/2)) F = 2.0*AJ/(AN-1.0)
        IF(I.GT.(N/2)) F = 2.0-2.0*AJ/(AN-1.0)
        B(I) = B(I)*F
390     CONTINUE
        RETURN
400     DO 490 I = 1,N
        AJ = FLOAT(I-1)
        F = 0.42-0.5*COS (TWOPI*AJ/(AN-1.0))+0.08*
       *  COS(4.0*PI*AJ/(AN-1.0))
        B(I) = B(I)*F
490     CONTINUE
        RETURN
        END
```

APPENDIX A.4 NINE TRACK TO SEVEN TRACK TAPE CONVERSION PROGRAM

```
      DIMENSION DAT(1024),IDAT(1024)
      FACTOR=(2.0**23)/250.0
      HTEST=2**23-1
      LTEST=-HTEST
      NFILES=6
      REWIND 2
      REWIND 4
      N=1C24
C
      DO 200 I=1,NFILES
      WRITE(6,11) I
11    FORMAT('1FILE',I4)
C
         DO 100 J=1,50
         READ(2,15,END=190,ERR=30) DAT
15       FORMAT(128A4)
         GO TO 50
30       WRITE(6,21)
21       FORMAT(60X,'READ ERROR')
50       WRITE(6,16) J
16       FORMAT(10X,'RECORD HAS BEEN READ',I4)
         IF(J.EQ.1) WRITE(6,17) DAT
17       FORMAT(1X,10F12.3)
C
            DO 80 K=1,1024
            IDAT(K)=IFIX(DAT(K)*FACTOR)
            IF(IDAT(K).GT.HTEST) WRITE(6,18) I,J,K
18          FORMAT(40X,'TOO LARGE FILE',I4,' RECORD',I4,
     *      ' ITEM',I4)
            IF(IDAT(K).LT.LTEST) WRITE(6,19) I,J,K
19          FORMAT(40X,'TOO SMALL FILE',I4,' RECORD',I4,
     *      ' ITEM',I4)
80          CONTINUE
C
         IF(J.EQ.1) WRITE(6,20) IDAT
20       FORMAT(1X,10I12)
         CALL MORF(IDAT,N)
         WRITE(4,25) IDAT
25       FORMAT(128(8A4))
100      CONTINUE
C
      WRITE(6,26)
26    FORMAT(5X,'ALL 50 RECORDS READ')
155   READ(2,15,END=190) DAT
      GO TO 155
190   WRITE(6,27)
27    FORMAT(2X,'END OF FILE')
      ENDFILE 4
200   CONTINUE
C
      STOP
      END
```

APPENDIX B.1 COMPUTER ANALYSIS AND MODIFICATION OF VOICED SPEECH

The 15 frame (384 msec.) segment of speech analyzed in this appendix is the "long e" sound (as in need) and is spoken by a woman. The process illustrated shows both direct reconstruction and reconstruction with the pitch reduced by a factor of 0.58 and the formant frequencies reduced by a factor of 0.88.

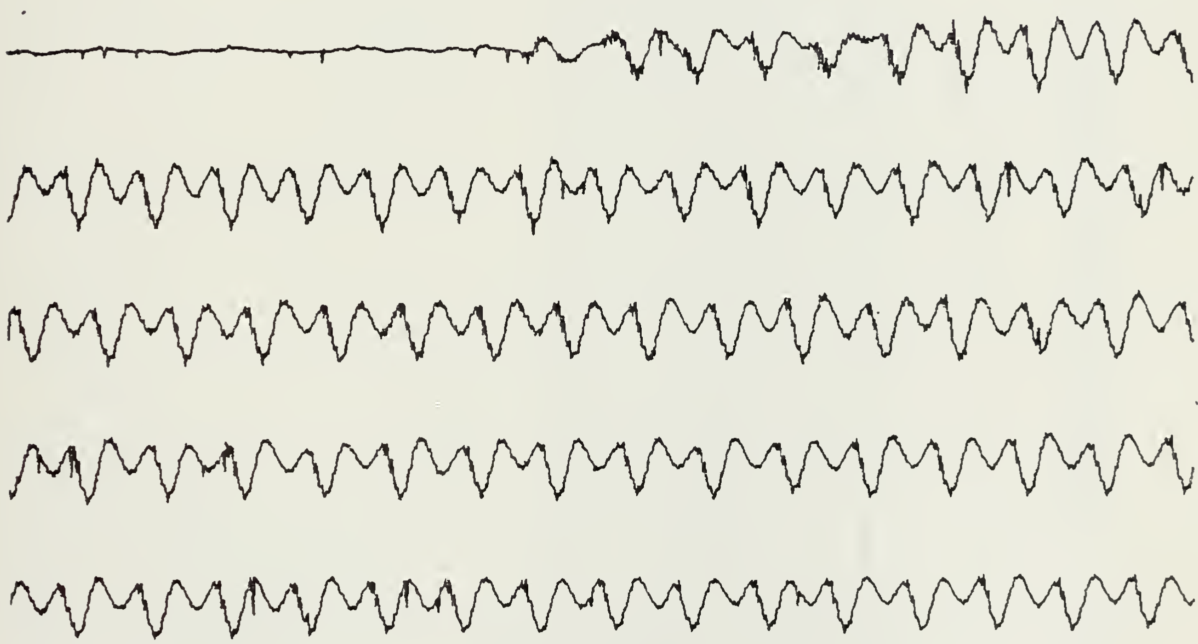


Figure B.1.1 WAVEFORM OF INPUT SPEECH

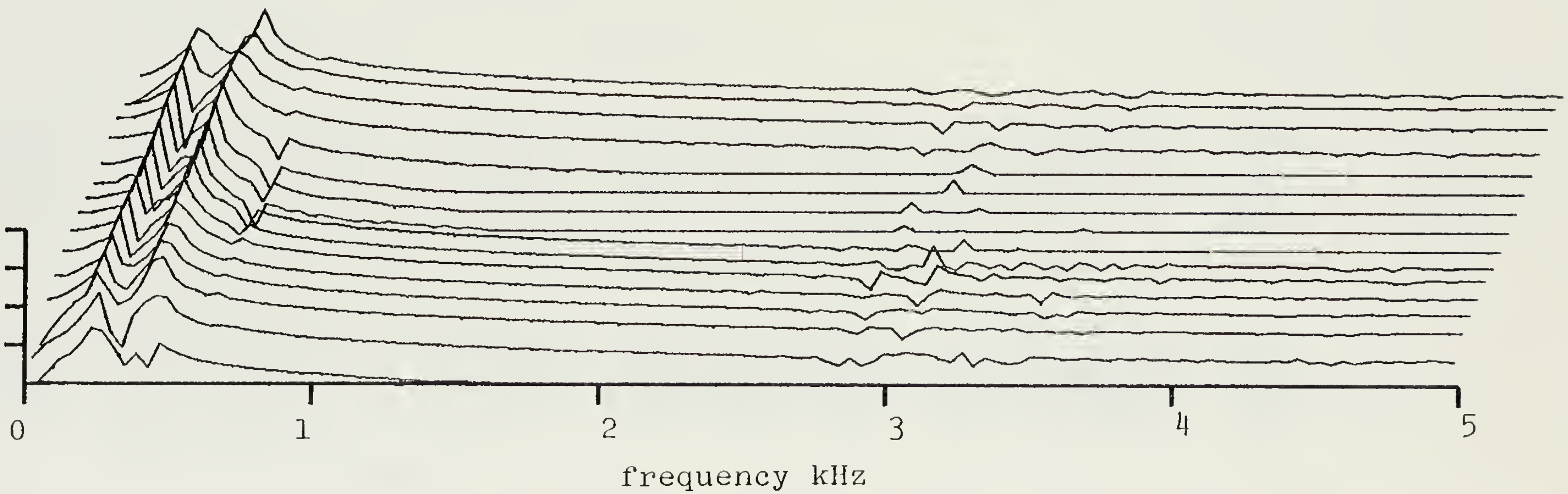


Figure B.1.2 LOGARITHMIC POWER SPECTRAL DENSITY OF INPUT SPEECH

```
FRAME     2

RMS VALUE OF SAMPLES   =          5.36754894

PREDICTOR COEFFICIENTS
          1           -0.34803427
          2           -0.23671557
          3           -0.55884212
          4           -0.13897109
          5            0.30623269
          6            0.04722288
          7            0.11424035
          8            0.11363018
          9            0.05982126
         10            0.03337795
         11           -0.18658018
         12            0.10059947

G IN =   3.93809

RMS VALUE OF ERROR   =          2.08631706

RATIO SAMPLE RMS TO ERROR RMS =          2.57273865

THIS FRAME IS VOICED

PITCH PERIOD IS  42
FORMANT  1 DUE TO POLES AT Z=  0.9509+-J*  0.1895 FORMANT FREQ=   313.0 BANDWIDTH=   49.2
FORMANT  2 DUE TO POLES AT Z=  0.5321+-J*  0.2559 FORMANT FREQ=   713.6 BANDWIDTH=  838.5
FORMANT  3 DUE TO POLES AT Z=  0.2907+-J*  0.7362 FORMANT FREQ=  1901.4 BANDWIDTH=  372.0
FORMANT  4 DUE TO POLES AT Z= -0.2239+-J*  0.8965 FORMANT FREQ=  2889.5 BANDWIDTH=  125.7
FORMANT  5 DUE TO POLES AT Z= -0.5320+-J*  0.6857 FORMANT FREQ=  3550.1 BANDWIDTH=  225.5
FORMANT  6 DUE TO POLES AT Z= -0.8438+-J*  0.2241 FORMANT FREQ=  4586.8 BANDWIDTH=  216.1

FORMANT FREQUENCY SCALE FACTOR =  0.8800   BANDWIDTH SCALE FACTOR =  0.6300
REAL POLE SCALE FACTOR =  1.0000  REAL POLE MAGNITUDE LIMIT =  0.9500  SAMPLE PERIOD = 0.000100

AFTER MODIFICATION
FORMANT  1 DUE TO POLES AT Z=  0.9752+-J*  0.1705 FORMANT FREQ=   275.4 BANDWIDTH=   16.0
FORMANT  2 DUE TO POLES AT Z=  0.6624+-J*  0.2758 FORMANT FREQ=   627.9 BANDWIDTH=  528.3
FORMANT  3 DUE TO POLES AT Z=  0.4284+-J*  0.7492 FORMANT FREQ=  1673.2 BANDWIDTH=  234.4
FORMANT  4 DUE TO POLES AT Z= -0.0255+-J*  0.9511 FORMANT FREQ=  2542.7 BANDWIDTH=   79.2
FORMANT  5 DUE TO POLES AT Z= -0.3495+-J*  0.8452 FORMANT FREQ=  3124.1 BANDWIDTH=  142.0
FORMANT  6 DUE TO POLES AT Z= -0.7548+-J*  0.5225 FORMANT FREQ=  4036.4 BANDWIDTH=  136.1
         RECON POLES  12

PITCH PERIOD AFTER MODIFICATION 73

RECONSTRUCTED POLYNOMIAL COEFFICIENTS
                         IMAGINARY TERMS SHOULD BE ZERO
      1        -1.87230908        0.0
      2         2.04516989        0.0
      3        -2.50993635        0.2220D-15
      4         2.31839986        0.1110D-14
      5        -1.96232965        0.2220D-14
      6         1.89773674        0.2887D-14
      7        -1.74324330        0.1332D-14
      8         1.66744029        0.2220D-15
      9        -1.34412913        0.0
     10         1.07349155        0.0
     11        -0.72689571       -0.5551D-16
     12         0.23989215        0.0

G OUT =  12.00667
```

Figure B.1.3(a) Processing Summary of Frame 2

```
FRAME     3

RMS VALUE OF SAMPLES   =          8.64423656
                                                           RECONSTRUCTED POLYNOMIAL COEFFICIENTS
PREDICTOR COEFFICIENTS                                                          IMAGINARY TERMS SHOULD BE ZERO
          1          -0.10656184                               1       -1.59896541      0.0
          2          -0.72204965                               2        1.58770708      0.0
          3          -1.34734794                               3       -2.81042781     -0.6661D-15
          4          -0.13047142                               4        2.50545591     -0.4441D-15
          5           1.15246868                               5       -2.05274495     -0.4441D-15
          6           1.19226933                               6        2.92916289      0.6661D-15
          7           0.14268279                               7       -2.13569256      0.4441D-15
          8          -0.55163610                               8        1.86288826      0.4441D-15
          9          -0.69186711                               9       -2.04557939     -0.2220D-15
         10          -0.03415674                              10        1.11747587     -0.5551D-16
         11           0.00061764                              11       -0.80281657      0.1388D-16
         12           0.28236032                              12        0.45104539      0.0

G IN =   7.36337                                           G CUT  =  21.04892

RMS VALUE OF ERROR   =          1.69222332

RATIO SAMPLE RMS TO ERROR RMS =          5.10815721

THIS FRAME IS VOICED

PITCH PERIOD IS  46
FORMANT  1 DUE TO POLES AT Z=  0.8922+-J*  0.1684 FORMANT FREQ=   300.2 BANDWIDTH=   171.1
FORMANT  2 DUE TO POLES AT Z=  0.9317+-J*  0.2636 FORMANT FREQ=   438.8 BANDWIDTH=    51.3
FORMANT  3 DUE TO POLES AT Z= -0.0395+-J*  0.7666 FORMANT FREQ=  2582.0 BANDWIDTH=   421.0
FORMANT  4 DUE TO POLES AT Z= -0.2765+-J*  0.9154 FORMANT FREQ=  2966.9 BANDWIDTH=    71.2
FORMANT  5 DUE TO POLES AT Z= -0.5586+-J*  0.6975 FORMANT FREQ=  3574.8 BANDWIDTH=   179.0
FORMANT  6 DUE TO POLES AT Z= -0.8859+-J*  0.2893 FORMANT FREQ=  4497.7 BANDWIDTH=   112.2

FORMANT FREQUENCY SCALE FACTOR =  0.8900   BANDWIDTH SCALE FACTOR =  0.6300
REAL POLE SCALE FACTOR =  1.0000   REAL POLE MAGNITUDE LIMIT =  0.9500  SAMPLE PERIOD = 0.000100

AFTER MODIFICATION
FORMANT  1 DUE TO POLES AT Z=  0.9217+-J*  0.1544 FORMANT FREQ=   264.1 BANDWIDTH=   107.8
FORMANT  2 DUE TO POLES AT Z=  0.9512+-J*  0.2354 FORMANT FREQ=   386.2 BANDWIDTH=    32.3
FORMANT  3 DUE TO POLES AT Z=  0.1208+-J*  0.8378 FORMANT FREQ=  2272.1 BANDWIDTH=   265.2
FORMANT  4 DUE TO POLES AT Z= -0.0677+-J*  0.9699 FORMANT FREQ=  2610.9 BANDWIDTH=    44.9
FORMANT  5 DUE TO POLES AT Z= -0.3677+-J*  0.8559 FORMANT FREQ=  3145.8 BANDWIDTH=   112.8
FORMANT  6 DUE TO POLES AT Z= -0.7588+-J*  0.5825 FORMANT FREQ=  3958.0 BANDWIDTH=    70.7
           RECON POLES  12

PITCH PERIOD AFTER MODIFICATION 80
```

Figure B.1.3(b) Processing Summary of Frame 3

```
FRAME    4

RMS VALUE OF SAMPLES   =          9.01073360

PREDICTOR COEFFICIENTS                               RECONSTRUCTED POLYNOMIAL COEFFICIENTS
          1        -0.23661983                                            IMAGINARY TERMS SHOULD BE ZERO
          2        -0.64776779                        1      -1.74887682      0.0
          3        -1.14240265                        2       1.81218555      0.0
          4        -0.05800193                        3      -2.67756438     -0.4441D-15
          5         1.00854492                        4       2.68510449      0.0
          6         0.81610531                        5      -2.18752132      0.0
          7         0.07864207                        6       2.82274383     -0.2220D-15
          8        -0.33500040                        7      -2.27572305     -0.6661D-15
          9        -0.46985948                        8       2.06317784     -0.4441D-15
         10         0.04502356                        9      -2.04035886      0.0
         11        -0.13275403                       10       1.25905941      0.5551D-16
         12         0.20858872                       11      -0.84527407     -0.5551D-16
                                                     12       0.37186913      0.0
G IN =   8.03218
                                                     G CUT =   28.71713
RMS VALUE OF ERROR   =          1.40273666

RATIO SAMPLE RMS TO ERROR RMS =          6.42368126

THIS FRAME IS VOICED

PITCH PERIOD IS  49
FORMANT  1 DUE TO POLES AT Z=  0.8053+-J*  0.1234 FORMANT FREQ=   242.1 BANDWIDTH=   326.2
FORMANT  2 DUE TO POLES AT Z=  0.9533+-J*  0.2502 FORMANT FREQ=   408.6 BANDWIDTH=    23.2
FORMANT  3 DUE TO POLES AT Z=  0.0425+-J*  0.7081 FORMANT FREQ=  2404.5 BANDWIDTH=   546.6
FORMANT  4 DUE TO POLES AT Z= -0.2590+-J*  0.9136 FORMANT FREQ=  2939.7 BANDWIDTH=    82.3
FORMANT  5 DUE TO POLES AT Z= -0.5341+-J*  0.7308 FORMANT FREQ=  3504.5 BANDWIDTH=   158.6
FORMANT  6 DUE TO POLES AT Z= -0.8896+-J*  0.2812 FORMANT FREQ=  4512.7 BANDWIDTH=   110.3

FORMANT FREQUENCY SCALE FACTOR =  0.8800    BANDWIDTH SCALE FACTOR =  0.6300
REAL POLE SCALE FACTOR =  1.0000   REAL POLE MAGNITUDE LIMIT =  0.9500  SAMPLE PERIOD = 0.000100

AFTER MODIFICATION
FORMANT  1 DUE TO POLES AT Z=  0.8710+-J*  0.1175 FORMANT FREQ=   213.0 BANDWIDTH=   205.5
FORMANT  2 DUE TO POLES AT Z=  0.9648+-J*  0.2217 FORMANT FREQ=   359.5 BANDWIDTH=    16.0
FORMANT  3 DUE TO POLES AT Z=  0.1925+-J*  0.7821 FORMANT FREQ=  2116.0 BANDWIDTH=   344.3
FORMANT  4 DUE TO POLES AT Z= -0.0528+-J*  0.9665 FORMANT FREQ=  2586.9 BANDWIDTH=    51.9
FORMANT  5 DUE TO POLES AT Z= -0.3369+-J*  0.8766 FORMANT FREQ=  3084.0 BANDWIDTH=    99.9
FORMANT  6 DUE TO POLES AT Z= -0.7641+-J*  0.5766 FORMANT FREQ=  3971.2 BANDWIDTH=    69.5
          RECON POLES  12

PITCH PERIOD AFTER MODIFICATION 85
```

Figure B.1.3(c) Processing Summary of Frame 4

```
FRAME     5

RMS VALUE OF SAMPLES   =           8.21990871

PREDICTOR COEFFICIENTS
                  1              -0.41879892
                  2              -0.37710579
                  3              -0.53962564
                  4               0.02388851
                  5               0.31795400
                  6               0.08458126
                  7               0.01706400
                  8               0.25840998
                  9               0.02474533
                 10               0.02720441
                 11              -0.22294211
                 12               0.01921266

G IN =    4.66012

RECONSTRUCTED POLYNOMIAL COEFFICIENTS
                                 IMAGINARY TERMS SHOULD BE ZERO
       1         -1.65272460          0.0
       2          1.69922547          0.0
       3         -2.23514145          0.0
       4          2.06739751          0.2220D-15
       5         -1.79939258          0.4667D-15
       6          1.79403631          0.8341D-16
       7         -1.64661295         -0.1072D-15
       8          1.49577653         -0.2686D-15
       9         -0.97841555         -0.4338D-16
      10          0.74783406          0.5893D-17
      11         -0.41767265          0.0
      12          0.03131921          0.0

G OUT =    9.46694

RMS VALUE OF ERROR   =           2.28235340

RATIO SAMPLE RMS TO ERROR RMS =            3.60150528

THIS FRAME IS VOICED

PITCH PERIOD IS   49
FORMANT  1 DUE TO POLES AT Z=   0.9478+-J*   0.2468 FORMANT FREQ=   405.5 BANDWIDTH=    33.2
FORMANT  2 DUE TO POLES AT Z=   0.3617+-J*   0.7338 FORMANT FREQ=  1771.1 BANDWIDTH=   319.6
FORMANT  3 DUE TO POLES AT Z=  -0.2394+-J*   0.8641 FORMANT FREQ=  2930.2 BANDWIDTH=   173.6
FORMANT  4 DUE TO POLES AT Z=  -0.4746+-J*   0.6726 FORMANT FREQ=  3478.0 BANDWIDTH=   309.7
FORMANT  5 DUE TO POLES AT Z=  -0.8352+-J*   0.2808 FORMANT FREQ=  4483.9 BANDWIDTH=   201.4
REAL POLE NUMBER    1 AT Z =   0.0872   0.0
REAL POLE NUMBER    2 AT Z =   0.8111   0.0

FORMANT FREQUENCY SCALE FACTOR =   0.8800    BANDWIDTH SCALE FACTOR =   0.6300
REAL POLE SCALE FACTOR =   1.0000   REAL POLE MAGNITUDE LIMIT =   0.9500   SAMPLE PERIOD = 0.000100

AFTER MODIFICATION
FORMANT  1 DUE TO POLES AT Z=   0.9652+-J*   0.2201 FORMANT FREQ=   356.8 BANDWIDTH=    16.0
FORMANT  2 DUE TO POLES AT Z=   0.4914+-J*   0.7314 FORMANT FREQ=  1558.6 BANDWIDTH=   201.4
FORMANT  3 DUE TO POLES AT Z=  -0.0461+-J*   0.9325 FORMANT FREQ=  2578.6 BANDWIDTH=   109.4
FORMANT  4 DUE TO POLES AT Z=  -0.3052+-J*   0.8303 FORMANT FREQ=  3060.7 BANDWIDTH=   195.1
FORMANT  5 DUE TO POLES AT Z=  -0.7281+-J*   0.5679 FORMANT FREQ=  3945.8 BANDWIDTH=   126.9
REAL POLE NUMBER    1 AT Z =   0.0872   0.0
REAL POLE NUMBER    2 AT Z =   0.8111   0.0
          RECON POLES  12

PITCH PERIOD AFTER MODIFICATION 85
```

Figure B.1.3(d) Processing Summary of Frame 5

```
FRAME     6

RMS VALUE OF SAMPLES   =          7.95691109

PREDICTOR COEFFICIENTS                                           RECONSTRUCTED POLYNOMIAL COEFFICIENTS
                                                                                       IMAGINARY TERMS SHOULD BE ZERO
          1           -0.42204028                                     1        -1.46354312        0.0
          2           -0.2316[illegible]                              2         1.30681951        0.0
          3           -0.599[illegible]                               3        -1.70020656       -0.2220D-15
          4           -0.0053[illegible]                              4         1.31279084       -0.7585D-16
          5            0.26102853                                     5        -0.98064812        0.3941D-15
          6            0.13934628                                     6         0.92359972        0.5457D-15
          7            0.05802783                                     7        -0.68438104       -0.4644D-15
          8            0.17145860                                     8         0.58109201       -0.3550D-15
          9            0.12297350                                     9        -0.23554245        0.8560D-16
         10           -0.06666410                                    10         0.28869809        0.9623D-16
         11           -0.13026506                                    11        -0.19441829        0.1129D-16
         12           -0.01715783                                    12        -0.03521242        0.0

G IN =    4.54883                                                G CUT =    8.39999

RMS VALUE OF ERROR   =          2.58917999

RATIO SAMPLE RMS TO ERROR RMS  =          3.07313919

THIS FRAME IS VOICED

PITCH PERIOD IS   49
FORMANT  1 DUE TO POLES AT Z=  0.9459+-J*  0.2394 FORMANT FREQ=   394.4 BANDWIDTH=   39.1
FORMANT  2 DUE TO POLES AT Z=  0.3474+-J*  0.7251 FORMANT FREQ=  1788.9 BANDWIDTH=  347.2
FORMANT  3 DUE TO POLES AT Z= -0.2382+-J*  0.8794 FORMANT FREQ=  2920.9 BANDWIDTH=  148.1
FORMANT  4 DUE TO POLES AT Z= -0.4965+-J*  0.5715 FORMANT FREQ=  3638.4 BANDWIDTH=  442.9
FORMANT  5 DUE TO POLES AT Z= -0.6769+-J*  0.2146 FORMANT FREQ=  4511.5 BANDWIDTH=  545.0
REAL POLE NUMBER   1 AT Z = -0.1447   0.0
REAL POLE NUMBER   2 AT Z =  0.8032   0.0

FORMANT FREQUENCY SCALE FACTOR =  0.8800    BANDWIDTH SCALE FACTOR =  0.6300
REAL POLE SCALE FACTOR =  1.0000  REAL POLE MAGNITUDE LIMIT =  0.9500  SAMPLE PERIOD = 0.000100

AFTER MODIFICATION
FORMANT  1 DUE TO POLES AT Z=  0.9665+-J*  0.2142 FORMANT FREQ=   347.1 BANDWIDTH=   16.0
FORMANT  2 DUE TO POLES AT Z=  0.4799+-J*  0.7282 FORMANT FREQ=  1574.2 BANDWIDTH=  218.7
FORMANT  3 DUE TO POLES AT Z= -0.0417+-J*  0.9421 FORMANT FREQ=  2570.4 BANDWIDTH=   93.3
FORMANT  4 DUE TO POLES AT Z= -0.3581+-J*  0.7589 FORMANT FREQ=  3201.8 BANDWIDTH=  279.1
FORMANT  5 DUE TO POLES AT Z= -0.6430+-J*  0.4859 FORMANT FREQ=  3970.1 BANDWIDTH=  343.3
REAL POLE NUMBER   1 AT Z = -0.1447   0.0
REAL POLE NUMBER   2 AT Z =  0.8032   0.0
        RECON POLES  12

PITCH PERIOD AFTER MODIFICATION 85
```

Figure B.1.3(e) Processing Summary of Frame 6

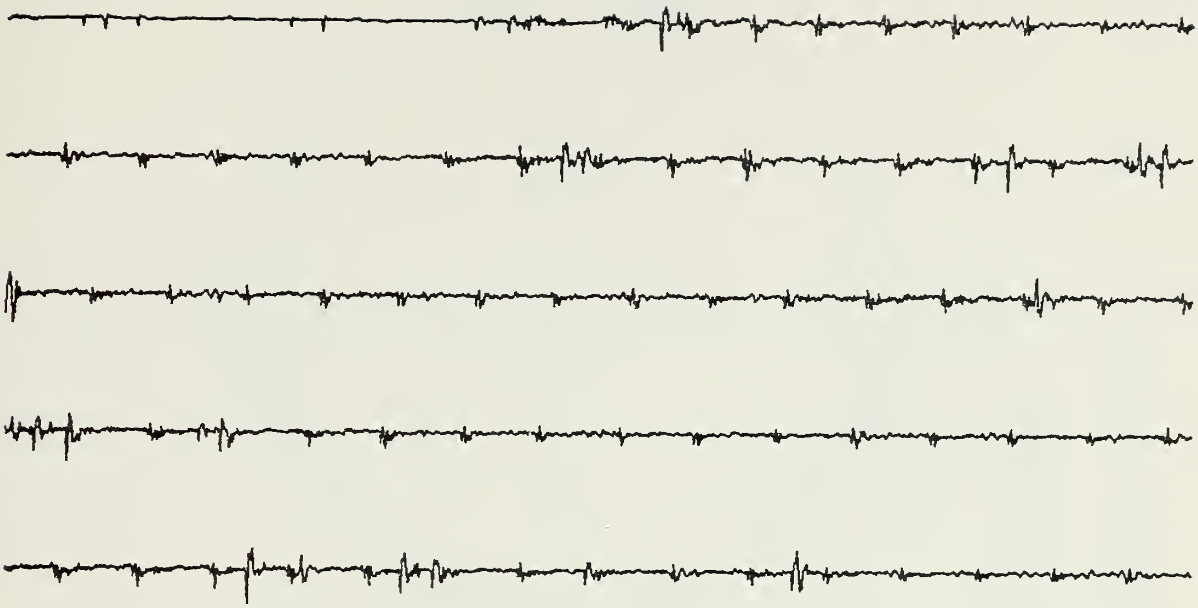


Figure B.1.4 WAVEFORM OF ERROR SIGNAL

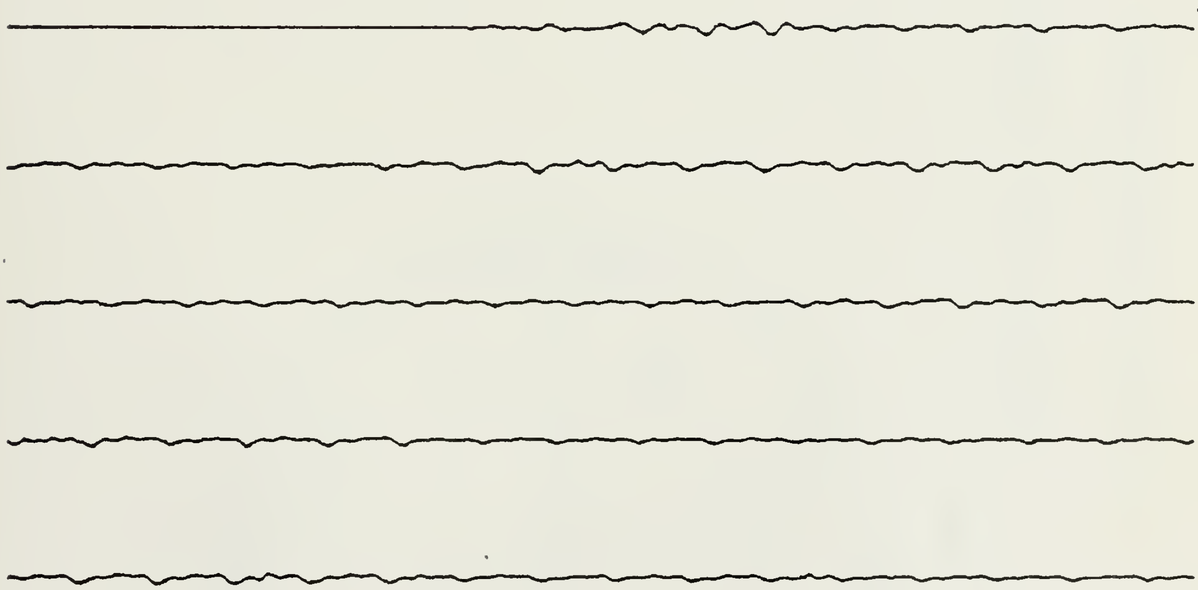


Figure B.1.5 WAVEFORM OF FILTERED ERROR SIGNAL

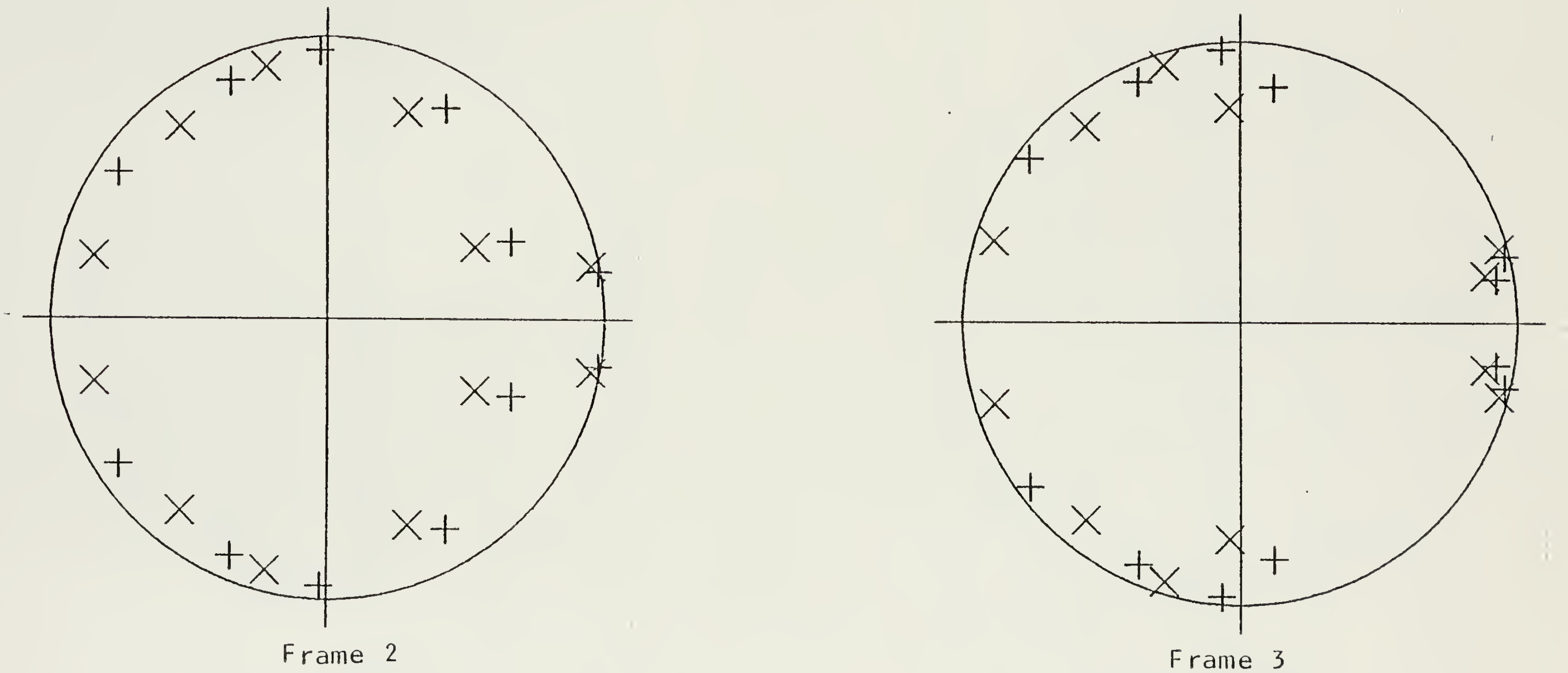
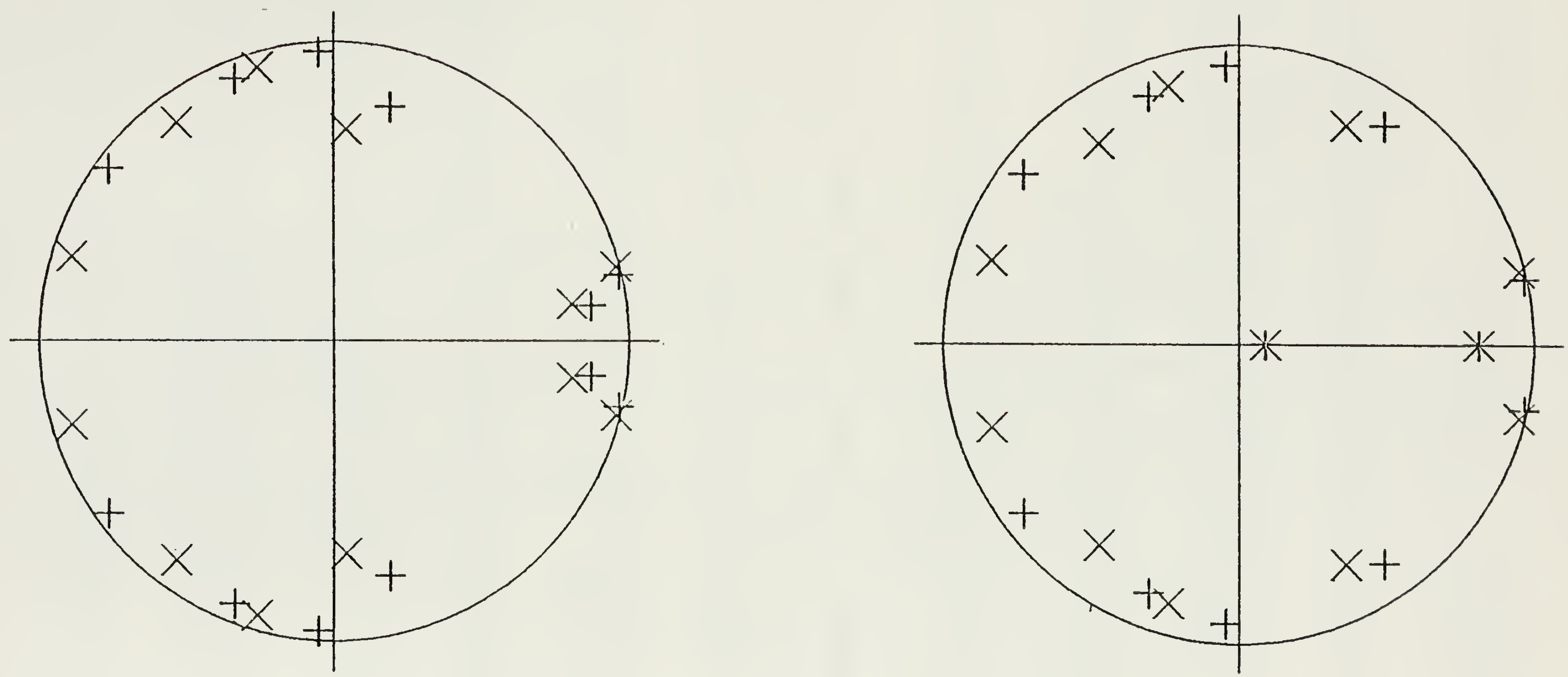


Figure B.1.6(a) VOCAL TRAÇT POLE LOCATIONS

X - Before Modification + - After Modification



Frame 4

Frame 5

Figure B.1.6(b) VOCAL TRACT POLE LOCATIONS

X - Before Modification + - After Modification

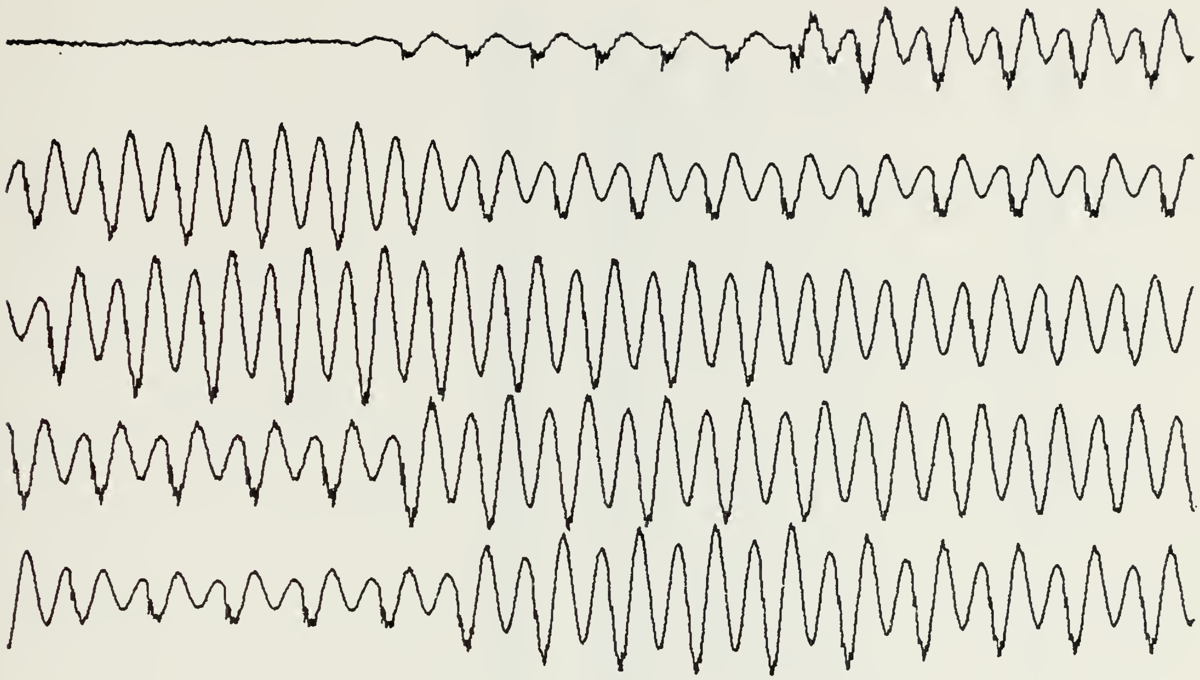


Figure B.1.7 WAVEFORM OF UNMODIFIED OUTPUT SPEECH

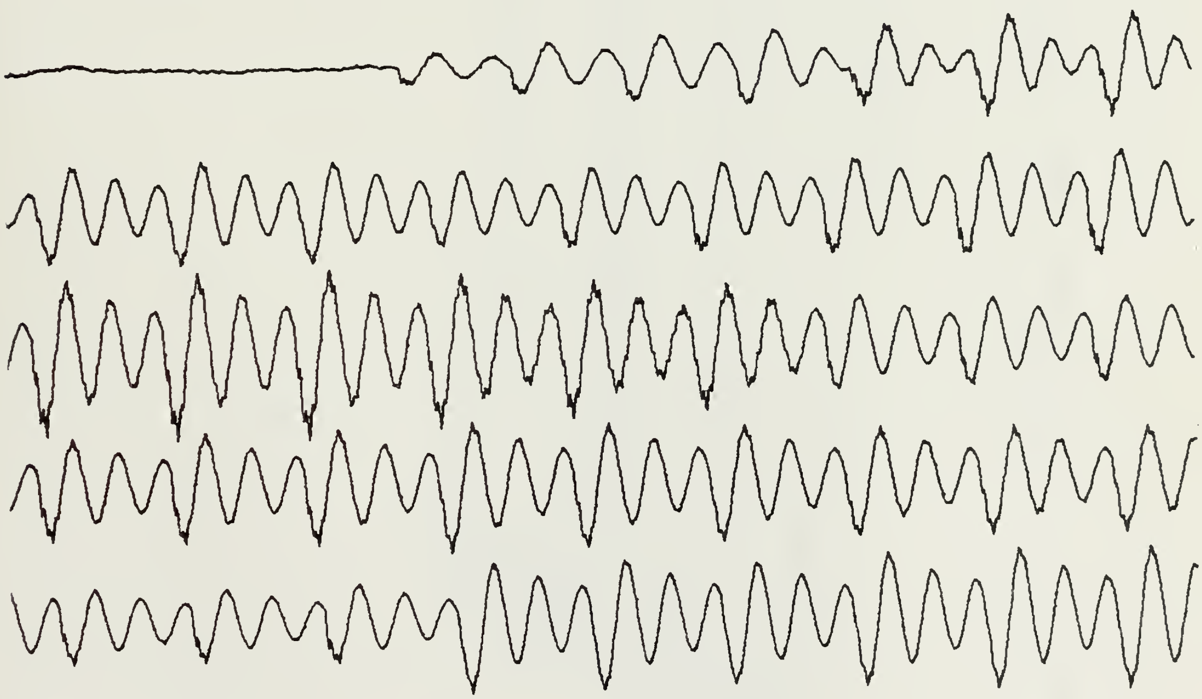


Figure B.1.8 WAVEFORM OF MODIFIED OUTPUT SPEECH

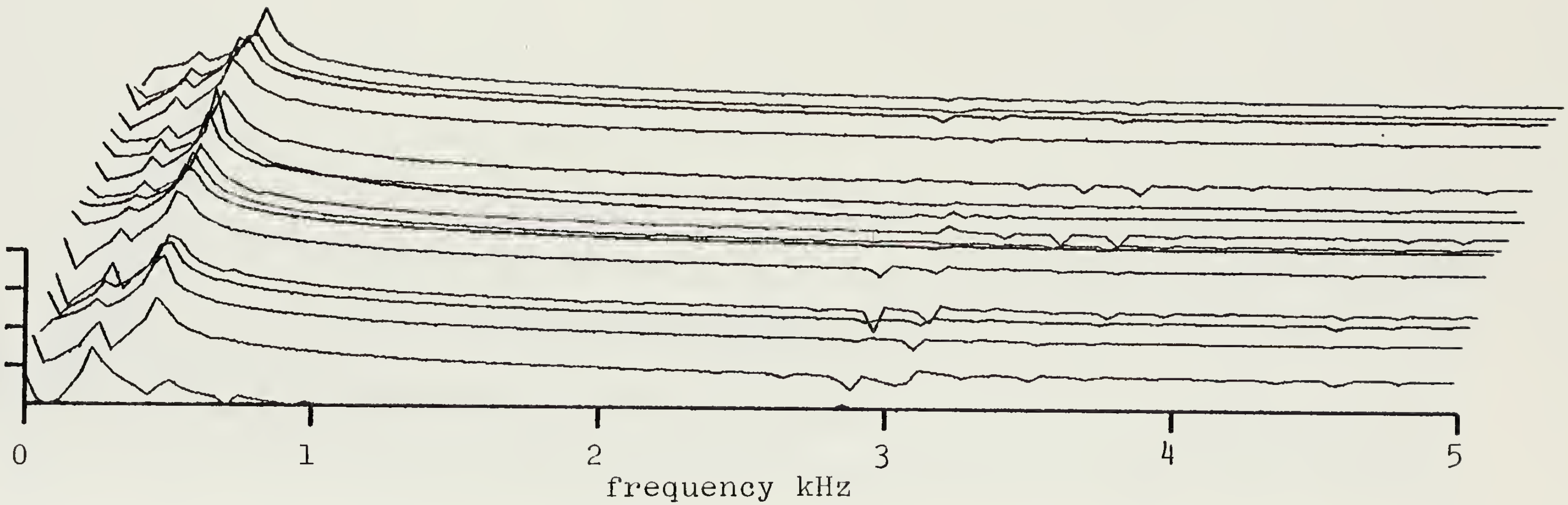


Figure B.1.9 LOGARITHMIC POWER SPECTRAL DENSITY OF UNMODIFIED OUTPUT SPEECH

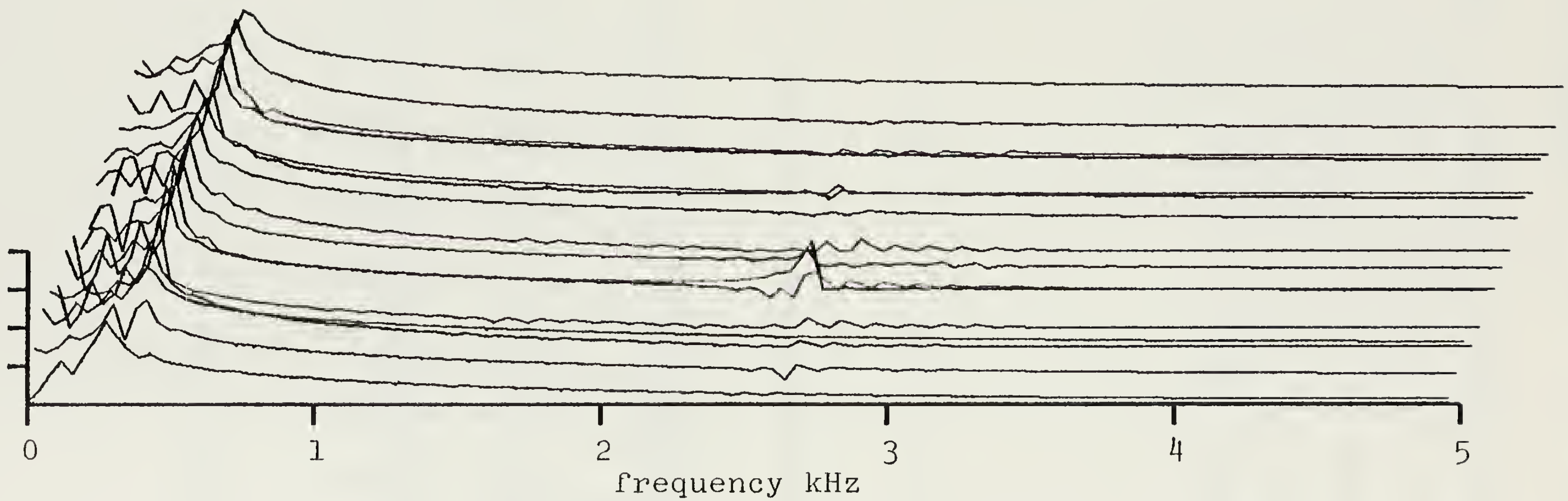


Figure B.1.10 LOGARITHMIC POWER SPECTRAL DENSITY OF MODIFIED OUTPUT SPEECH

APPENDIX B.2 COMPUTER ANALYSIS AND MODIFICATION OF UNVOICED SPEECH

The 15 frame (384 msec.) segment of speech analyzed in this appendix is the "sa" sound (begining of salt) and is spoken by a woman. The process illustrated shows both direct reconstruction and reconstruction with the pitch reduced by a factor of 0.58 and the formant frequencies reduced by a factor of 0.88.

Figure B.2.1 WAVEFORM OF INPUT SPEECH

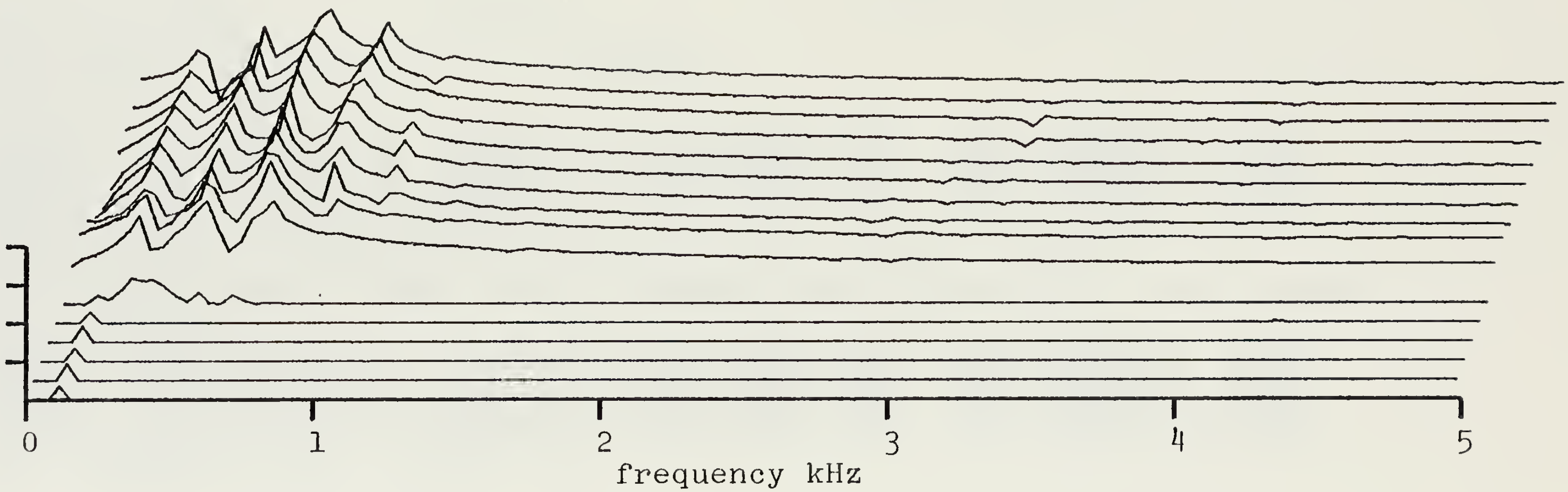


Figure B.2.2 LOGARITHMIC POWER SPECTRAL DENSITY OF INPUT SPEECH

```
FRAME     2

RMS VALUE OF SAMPLES   =          1.29037189

PREDICTOR COEFFICIENTS                                          RECONSTRUCTED POLYNOMIAL COEFFICIENTS
                                                                                 IMAGINARY TERMS SHOULD BE ZERO
          1          -0.12835950                                1        -1.32524467        0.0
          2          -0.17267976                                2         1.20391764        0.0
          3          -0.12716526                                3        -1.12342042       -0.6661D-15
          4          -0.23059702                                4         0.75638282       -0.1824D-15
          5          -0.00626768                                5        -0.74483762        0.1350D-15
          6          -0.03280819                                6         0.64663675        0.4849D-15
          7           0.01079201                                7        -0.62569162        0.1956D-15
          8           0.08192515                                8         0.54178444       -0.1151D-16
          9          -0.04018458                                9        -0.31577952       -0.1151D-15
         10           0.08133036                               10         0.12936728        0.2712D-17
         11           0.00123771                               11         0.06316667        0.1192D-16
         12          -0.01679996                               12        -0.04070372       -0.2549D-17

G IN =   2.37969                                                G OUT =   6.03541

RMS VALUE OF ERROR   =          1.03036118

RATIO SAMPLE RMS TO ERROR RMS =           1.25234890

THIS FRAME IS UNVOICED

FORMANT  1 DUE TO POLES AT Z=  0.8002+-J*  0.1395 FORMANT FREQ=   274.7 BANDWIDTH=   330.9
FORMANT  2 DUE TO POLES AT Z=  0.3817+-J*  0.6213 FORMANT FREQ=  1623.2 BANDWIDTH=   502.6
FORMANT  3 DUE TO POLES AT Z=  0.0280+-J*  0.7762 FORMANT FREQ=  2442.7 BANDWIDTH=   402.2
FORMANT  4 DUE TO POLES AT Z= -0.3940+-J*  0.6996 FORMANT FREQ=  3316.3 BANDWIDTH=   349.3
FORMANT  5 DUE TO POLES AT Z= -0.7815+-J*  0.2441 FORMANT FREQ=  4518.2 BANDWIDTH=   318.4
REAL POLE NUMBER    1 AT Z = -0.3999   0.0
REAL POLE NUMBER    2 AT Z =  0.4593   0.0

FORMANT FREQUENCY SCALE FACTOR =  0.8800   BANDWIDTH SCALE FACTOR =  0.6300
REAL POLE SCALE FACTOR =  1.0000  REAL POLE MAGNITUDE LIMIT =  0.9500  SAMPLE PERIOD = 0.000100

AFTER MODIFICATION
FORMANT  1 DUE TO POLES AT Z=  0.8671+-J*  0.1327 FORMANT FREQ=   241.7 BANDWIDTH=   208.5
FORMANT  2 DUE TO POLES AT Z=  0.5111+-J*  0.6407 FORMANT FREQ=  1428.4 BANDWIDTH=   316.6
FORMANT  3 DUE TO POLES AT Z=  0.1863+-J*  0.8322 FORMANT FREQ=  2149.5 BANDWIDTH=   253.4
FORMANT  4 DUE TO POLES AT Z= -0.2263+-J*  0.8410 FORMANT FREQ=  2918.3 BANDWIDTH=   220.0
FORMANT  5 DUE TO POLES AT Z= -0.7053+-J*  0.5289 FORMANT FREQ=  3976.0 BANDWIDTH=   200.6
REAL POLE NUMBER    1 AT Z = -0.3999   0.0
REAL POLE NUMBER    2 AT Z =  0.4593   0.0
          RECON POLES  12
```

Figure B.2.3(a) Processing Summary of Frame 2

```
FRAME    3

RMS VALUE OF SAMPLES  =          1.17910290

PRECICTOR COEFFICIENTS                                         RECONSTRUCTED POLYNOMIAL COEFFICIENTS
                                                                                      IMAGINARY TERMS SHOULD BE ZERO
            1        0.10897994                                      1        -1.26795423         0.0
            2       -0.85255343                                      2         0.77064886         0.0
            3       -0.35400796                                      3        -1.11166646        -0.4441D-15
            4        0.06470495                                      4         1.19483614        -0.4441D-15
            5        0.01876231                                      5        -1.14832072         0.0
            6        0.00943553                                      6         1.07585437         0.2220D-15
            7        0.04628577                                      7        -0.97424820        -0.2220D-15
            8        0.05911457                                      8         0.89147840        -0.4441D-15
            9       -0.03087259                                      9        -0.70581481        -0.8882D-15
           10       -0.10893565                                     10         0.51415969        -0.6106D-15
           11        0.01468784                                     11        -0.50656193        -0.3053D-15
           12        0.16223890                                     12         0.31815649         0.0

G IN =    7.24431                                              G OUT =   19.77582

RMS VALUE OF ERROR  =          0.64461362

RATIO SAMPLE RMS TO ERROR RMS =          1.82916164

THIS FRAME IS UNVOICED

FORMANT  1 DUE TO POLES AT Z=  0.9355+-J*  0.0893 FORMANT FREQ=   151.5 BANDWIDTH=   98.8
FORMANT  2 DUE TO POLES AT Z=  0.6390+-J*  0.4849 FORMANT FREQ=  1033.1 BANDWIDTH=  350.8
FORMANT  3 DUE TO POLES AT Z=  0.2283+-J*  0.7949 FORMANT FREQ=  2054.8 BANDWIDTH=  302.3
FORMANT  4 DUE TO POLES AT Z= -0.2988+-J*  0.7769 FORMANT FREQ=  3084.3 BANDWIDTH=  292.0
FORMANT  5 DUE TO POLES AT Z= -0.6838+-J*  0.5317 FORMANT FREQ=  3948.1 BANDWIDTH=  228.7
FORMANT  6 DUE TO POLES AT Z= -0.8749+-J*  0.1955 FORMANT FREQ=  4650.0 BANDWIDTH=  174.0

FORMANT FREQUENCY SCALE FACTOR =  0.8800    BANDWIDTH SCALE FACTOR =  0.6300
REAL POLE SCALE FACTOR =  1.0000  REAL POLE MAGNITUDE LIMIT =  0.9500  SAMPLE PERIOD = 0.000100

AFTER MODIFICATION
FORMANT  1 DUE TO POLES AT Z=  0.9583+-J*  0.0805 FORMANT FREQ=   133.3 BANDWIDTH=   62.3
FORMANT  2 DUE TO POLES AT Z=  0.7322+-J*  0.4705 FORMANT FREQ=   909.1 BANDWIDTH=  221.0
FORMANT  3 DUE TO POLES AT Z=  0.3736+-J*  0.8047 FORMANT FREQ=  1808.2 BANDWIDTH=  190.4
FORMANT  4 DUE TO POLES AT Z= -0.1195+-J*  0.8828 FORMANT FREQ=  2714.2 BANDWIDTH=  184.0
FORMANT  5 DUE TO POLES AT Z= -0.5249+-J*  0.7476 FORMANT FREQ=  3474.4 BANDWIDTH=  144.1
FORMANT  6 DUE TO POLES AT Z= -0.7856+-J*  0.5041 FORMANT FREQ=  4092.0 BANDWIDTH=  109.6
         RECON POLES  12
```

Figure B.2.3(b) Processing Summary of Frame 3

```
FRAME     4

RMS VALUE OF SAMPLES   =          1.23555470

PREDICTOR COEFFICIENTS                                                     RECONSTRUCTED POLYNOMIAL COEFFICIENTS
                                                                                             IMAGINARY TERMS SHOULD BE ZERO
            1          0.16449492                                     1          -0.86928492        0.0
            2         -0.91081697                                     2           0.10286410        0.0
            3         -0.31257844                                     3          -0.37505735       -0.2220D-15
            4         -0.03435435                                     4           0.21024421       -0.1874D-15
            5         -0.08130842                                     5          -0.31136055       -0.4487D-16
            6          0.14074278                                     6           0.24657936        0.2882D-15
            7          0.22526640                                     7          -0.02687662        0.3195D-15
            8          0.19467813                                     8           0.20465510       -0.3218D-16
            9         -0.06394994                                     9           0.11857971       -0.1486D-15
           10         -0.11490434                                    10          -0.19327977       -0.4338D-16
           11         -0.01946801                                    11           0.03640319        0.8487D-17
           12         -0.01048763                                    12          -0.04806215        0.5568D-18

G IN =    5.64007                                                 G CUT =  10.48170

RMS VALUE OF ERROR   =          0.75049871

RATIO SAMPLE RMS TO ERROR RMS =          1.64631081

THIS FRAME IS UNVOICED

FORMANT  1 DUE TO POLES AT Z=  0.8167+-J*  0.2211 FORMANT FREQ=   420.7 BANDWIDTH=   266.1
FORMANT  2 DUE TO POLES AT Z=  0.2365+-J*  0.7810 FORMANT FREQ=  2032.0 BANDWIDTH=   323.6
FORMANT  3 DUE TO POLES AT Z= -0.0455+-J*  0.2919 FORMANT FREQ=  2746.1 BANDWIDTH=  1940.8
FORMANT  4 DUE TO POLES AT Z= -0.3649+-J*  0.6398 FORMANT FREQ=  3324.9 BANDWIDTH=   486.8
FORMANT  5 DUE TO POLES AT Z= -0.8030+-J*  0.2818 FORMANT FREQ=  4462.9 BANDWIDTH=   256.7
REAL POLE NUMBER   1 AT Z = -0.7270   0.0
REAL POLE NUMBER   2 AT Z =  0.8830   0.0

FORMANT FREQUENCY SCALE FACTOR =  0.6800   BANDWIDTH SCALE FACTOR =  0.6300
REAL POLE SCALE FACTOR =  1.0000  REAL POLE MAGNITUDE LIMIT =  0.9500  SAMPLE PERIOD = 0.000100

AFTER MODIFICATION
FORMANT  1 DUE TO POLES AT Z=  0.8758+-J*  0.2075 FORMANT FREQ=   370.2 BANDWIDTH=   167.6
FORMANT  2 DUE TO POLES AT Z=  0.3805+-J*  0.7932 FORMANT FREQ=  1788.2 BANDWIDTH=   203.9
FORMANT  3 DUE TO POLES AT Z=  0.0243+-J*  0.4632 FORMANT FREQ=  2416.5 BANDWIDTH=  1222.7
FORMANT  4 DUE TO POLES AT Z= -0.2181+-J*  0.7954 FORMANT FREQ=  2925.9 BANDWIDTH=   306.7
FORMANT  5 DUE TO POLES AT Z= -0.7058+-J*  0.5638 FORMANT FREQ=  3927.3 BANDWIDTH=   161.7
REAL POLE NUMBER   1 AT Z = -0.7270   0.0
REAL POLE NUMBER   2 AT Z =  0.9830   0.0
          RECON POLES  12
```

Figure B.2.3(c) Processing Summary of Frame 4

```
FRAME     5

RMS VALUE OF SAMPLES  =            1.75883007

PREDICTOR COEFFICIENTS
          1            0.31991321
          2           -0.71884066
          3           -0.19736006
          4            0.09484917
          5           -0.29558820
          6            0.00791182
          7            0.17295140
          8            0.03932114
          9           -0.12524700
         10           -0.13390845
         11            0.06119598
         12            0.06722248

G IN =   3.41980

RMS VALUE OF ERROR  =            1.05495548

RATIO SAMPLE RMS TO ERROR RMS =          1.66720772

THIS FRAME IS UNVOICED

FORMANT  1 DUE TO POLES AT Z=  0.6219+-J*  0.4682 FORMANT FREQ=  1027.1 BANDWIDTH=   398.6
FORMANT  2 DUE TO POLES AT Z=  0.2190+-J*  0.8466 FORMANT FREQ=  2097.1 BANDWIDTH=   213.5
FORMANT  3 DUE TO POLES AT Z= -0.3304+-J*  0.6651 FORMANT FREQ=  3232.9 BANDWIDTH=   473.6
FORMANT  4 DUE TO POLES AT Z= -0.7977+-J*  0.4173 FORMANT FREQ=  4232.8 BANDWIDTH=   167.2
FORMANT  5 DUE TO POLES AT Z= -0.6913+-J*  0.1031 FORMANT FREQ=  4764.5 BANDWIDTH=   570.0
REAL POLE NUMBER   1 AT Z =  0.7492   0.0
REAL POLE NUMBER   2 AT Z =  0.8880   0.0

FORMANT FREQUENCY SCALE FACTOR =  0.8800   BANDWIDTH SCALE FACTOR =  0.6300
REAL POLE SCALE FACTOR =  1.0000  REAL POLE MAGNITUDE LIMIT =  0.9500  SAMPLE PERIOD = 0.000100

AFTER MODIFICATION
FORMANT  1 DUE TO POLES AT Z=  0.7200+-J*  0.4594 FORMANT FREQ=   903.9 BANDWIDTH=   251.1
FORMANT  2 DUE TO POLES AT Z=  0.3674+-J*  0.8423 FORMANT FREQ=  1845.4 BANDWIDTH=   134.5
FORMANT  3 DUE TO POLES AT Z= -0.1787+-J*  0.8096 FORMANT FREQ=  2845.8 BANDWIDTH=   298.4
FORMANT  4 DUE TO POLES AT Z= -0.6513+-J*  0.6722 FORMANT FREQ=  3724.9 BANDWIDTH=   105.3
FORMANT  5 DUE TO POLES AT Z= -0.6975+-J*  0.3876 FORMANT FREQ=  4192.7 BANDWIDTH=   359.1
REAL POLE NUMBER   1 AT Z =  0.7492   0.0
REAL POLE NUMBER   2 AT Z =  0.8880   0.0
          RECON POLES  12

RECONSTRUCTED POLYNOMIAL COEFFICIENTS
                              IMAGINARY TERMS SHOULD BE ZERO
     1         -0.75674117           0.0
     2          0.19341960           0.0
     3         -0.45038067          -0.1388D-16
     4          0.61465852           0.4668D-15
     5         -0.67045574           0.9639D-16
     6          0.38643425          -0.4294D-15
     7         -0.20071594          -0.3247D-15
     8          0.25026288           0.1520D-15
     9         -0.14704765           0.9691D-16
    10         -0.02470761           0.2272D-16
    11         -0.17401031          -0.9232D-17
    12          0.15711490           0.0

G CUT =   5.62331
```

Figure 2.3(d) Processing Summary of Frame 5

```
FRAME    6

RMS VALUE OF SAMPLES  =          1.67211628

PREDICTOR COEFFICIENTS
                1          0.42480206
                2         -0.41482103
                3         -0.23569275
                4         -0.05882226
                5         -0.27487868
                6          0.12121958
                7          0.18722564
                8         -0.05819549
                9         -0.11492920
               10         -0.07753271
               11         -0.12846768
               12         -0.02238397

G IN =    3.36109

RMS VALUE OF ERROR   =          1.24257047

RATIO SAMPLE RMS TO ERROR RMS =          1.34525776

THIS FRAME IS UNVOICED

FORMANT  1 DUE TO POLES AT Z=  0.7362+-J*  0.4247 FORMANT FREQ=   832.7 BANDWIDTH=   258.9
FORMANT  2 DUE TO POLES AT Z=  0.1910+-J*  0.7645 FORMANT FREQ=  2110.3 BANDWIDTH=   379.2
FORMANT  3 DUE TO POLES AT Z= -0.0418+-J*  0.7139 FORMANT FREQ=  2593.1 BANDWIDTH=   533.7
FORMANT  4 DUE TO POLES AT Z= -0.6837+-J*  0.5391 FORMANT FREQ=  3937.4 BANDWIDTH=   220.4
FORMANT  5 DUE TO POLES AT Z= -0.7941+-J*  0.2803 FORMANT FREQ=  4460.0 BANDWIDTH=   273.5
REAL POLE NUMBER    1 AT Z = -0.1909   0.0
REAL POLE NUMBER    2 AT Z =  0.9509   0.0

FORMANT FREQUENCY SCALE FACTOR =  0.8800   BANDWIDTH SCALE FACTOR =  0.6300
REAL POLE SCALE FACTOR =  1.0000   REAL POLE MAGNITUDE LIMIT =  0.9500  SAMPLE PERIOD = 0.000100

AFTER MODIFICATION
FORMANT  1 DUE TO POLES AT Z=  0.8086+-J*  0.4011 FORMANT FREQ=   732.8 BANDWIDTH=   163.1
FORMANT  2 DUE TO POLES AT Z=  0.3383+-J*  0.7913 FORMANT FREQ=  1857.0 BANDWIDTH=   238.9
FORMANT  3 DUE TO POLES AT Z=  0.1106+-J*  0.8020 FORMANT FREQ=  2281.9 BANDWIDTH=   336.2
FORMANT  4 DUE TO POLES AT Z= -0.5222+-J*  0.7531 FORMANT FREQ=  3464.9 BANDWIDTH=   138.8
FORMANT  5 DUE TO POLES AT Z= -0.7003+-J*  0.5612 FORMANT FREQ=  3924.8 BANDWIDTH=   172.3
REAL POLE NUMBER    1 AT Z = -0.1909   0.0
REAL POLE NUMBER    2 AT Z =  0.9500   0.0
        RECON POLES  12

RECONSTRUCTED POLYNOMIAL COEFFICIENTS
                     IMAGINARY TERMS SHOULD BE ZERO
    1       -0.82910281         0.0
    2        0.64258707         0.0
    3       -0.88929738        -0.4441D-15
    4        0.89355681        -0.5511D-15
    5       -1.09729040         0.5327D-15
    6        0.85180693         0.9403D-15
    7       -0.44287412         0.9024D-16
    8        0.46742224        -0.3895D-15
    9       -0.42176654        -0.1774D-15
   10        0.22974490        -0.6258D-17
   11       -0.19073813        -0.1053D-16
   12       -0.04851077        -0.2517D-17

G CUT =    6.02635
```

Figure B.2.3(e) Processing Summary of Frame 6

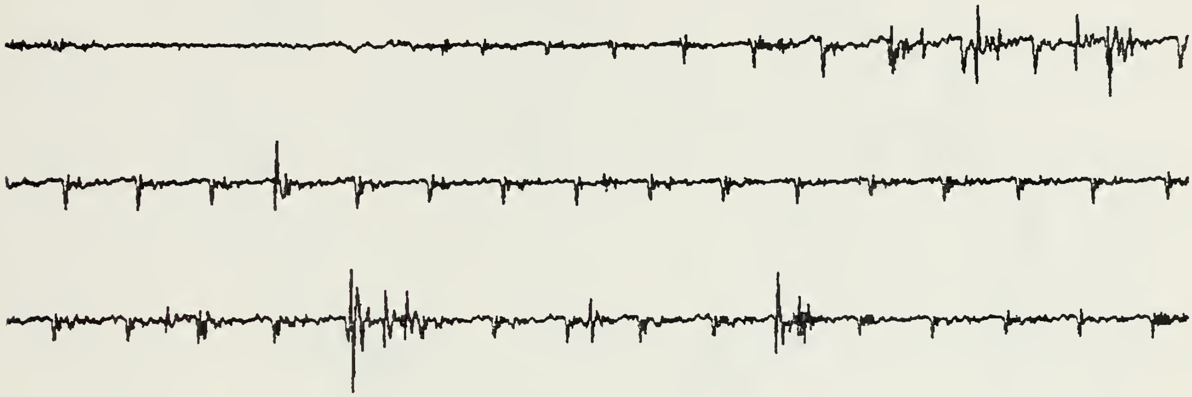


Figure B.2.4 WAVEFORM OF ERROR SIGNAL

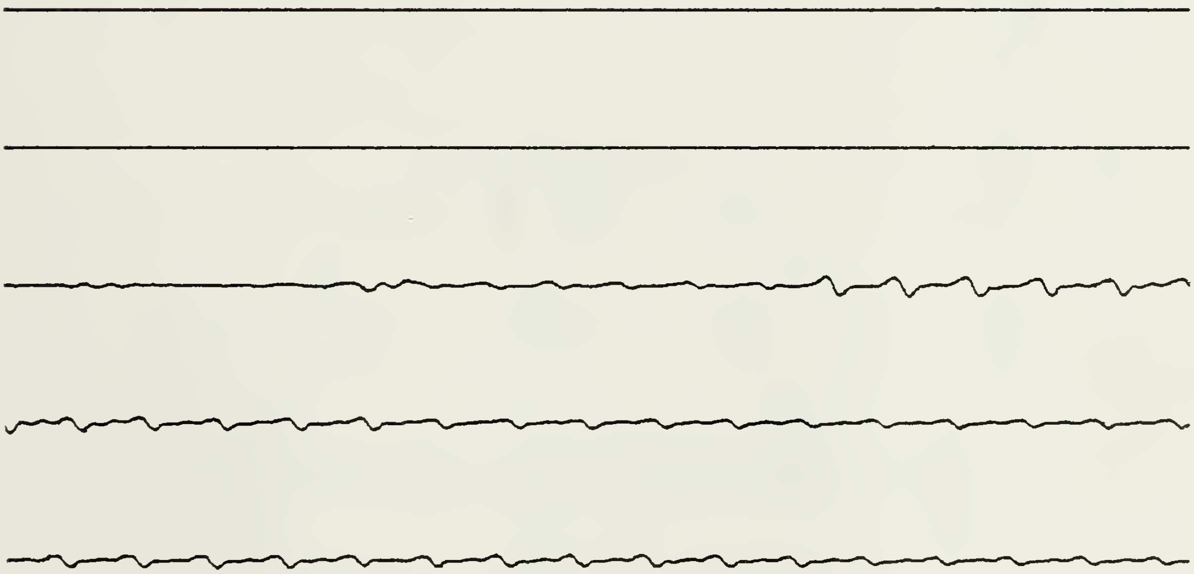
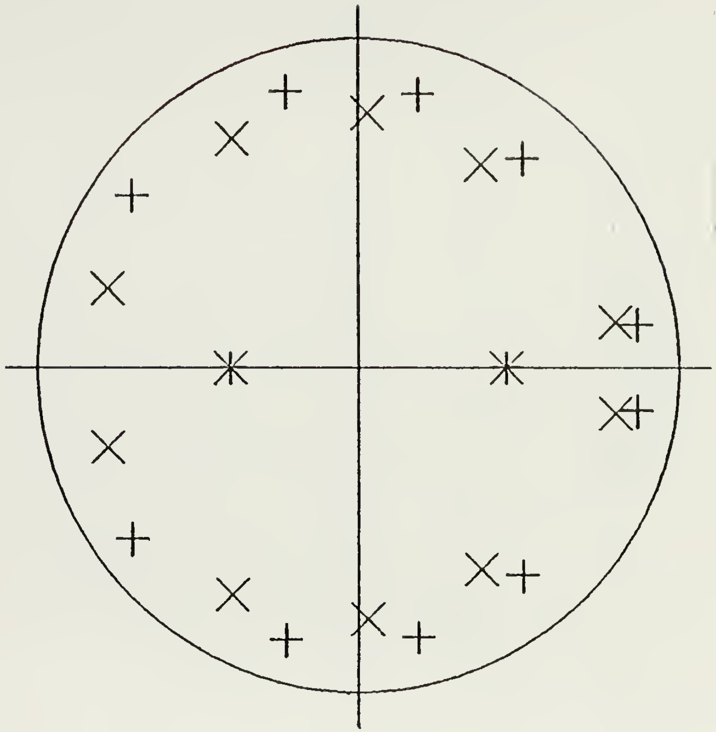
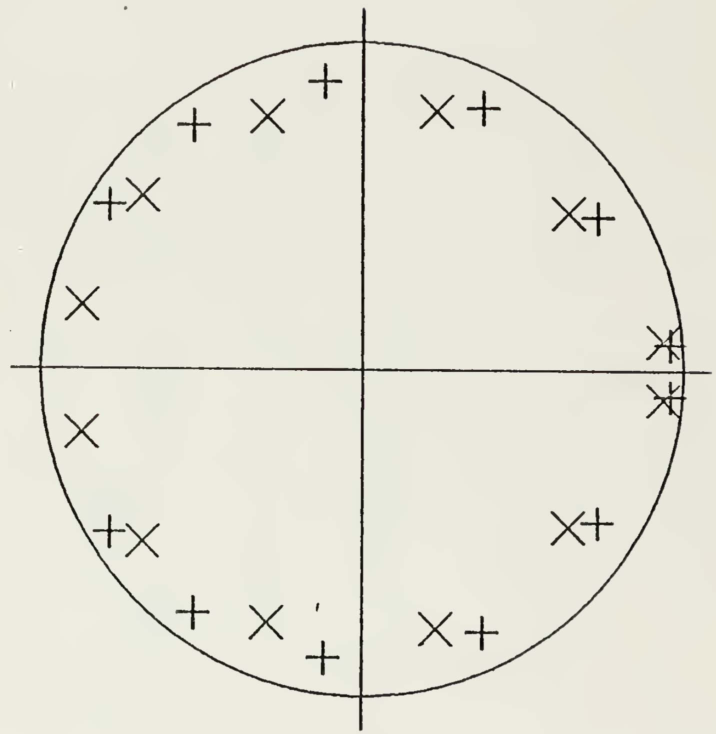


Figure B.2.5 WAVEFORM OF FILTERED ERROR SIGNAL



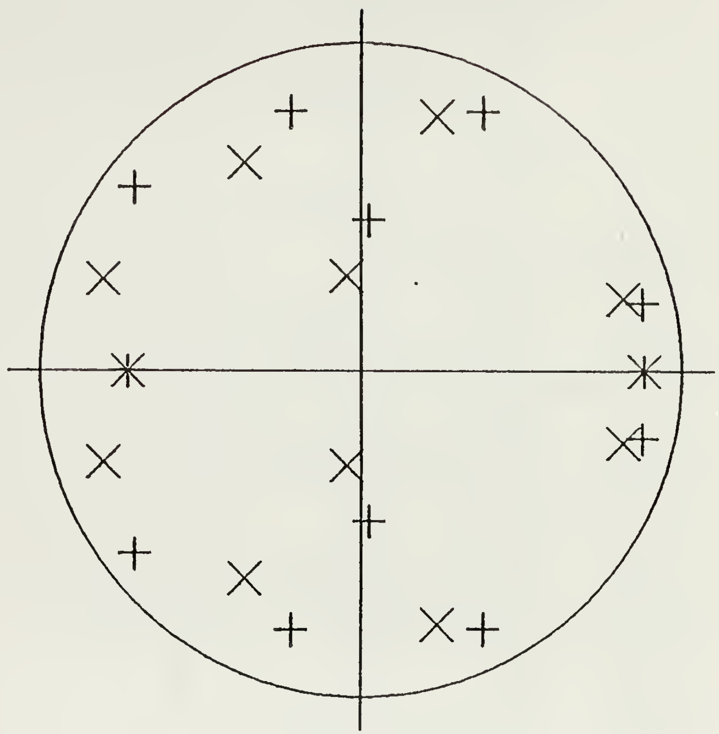
Frame 2



Frame 3

Figure B.2.6(a) VOCAL TRACT POLE LOCATIONS

X - Before Modification + - After Modification



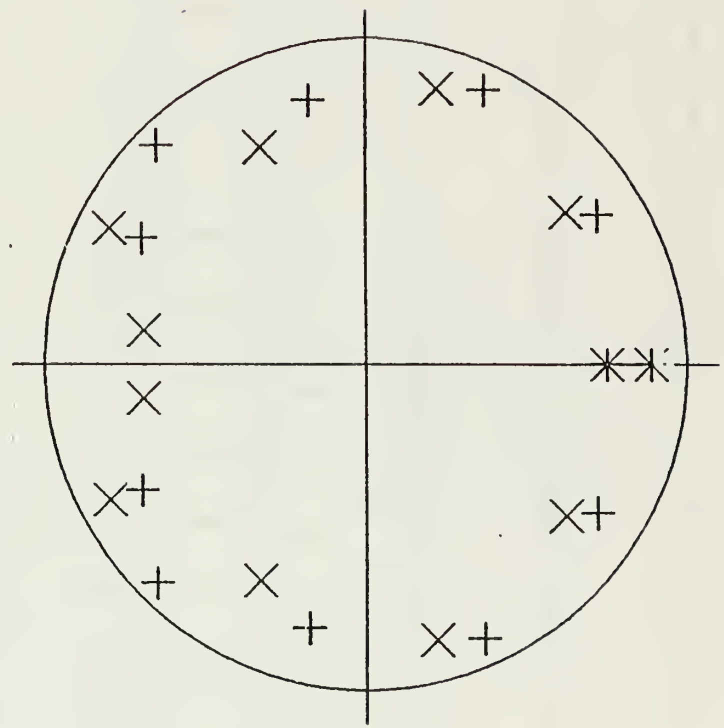


Figure B.2.6(b) VOCAL TRACT POLE LOCATIONS

X - Before Modification + - After Modification

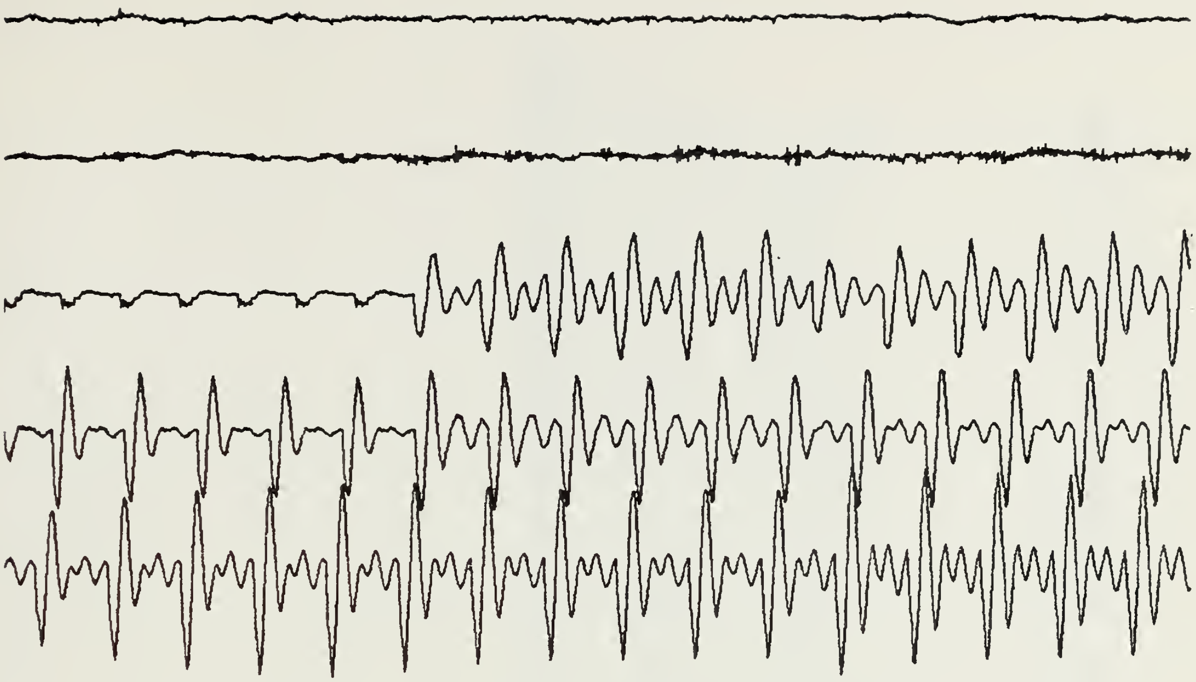


Figure B.2.7 WAVEFORM OF UNMODIFIED OUTPUT SPEECH

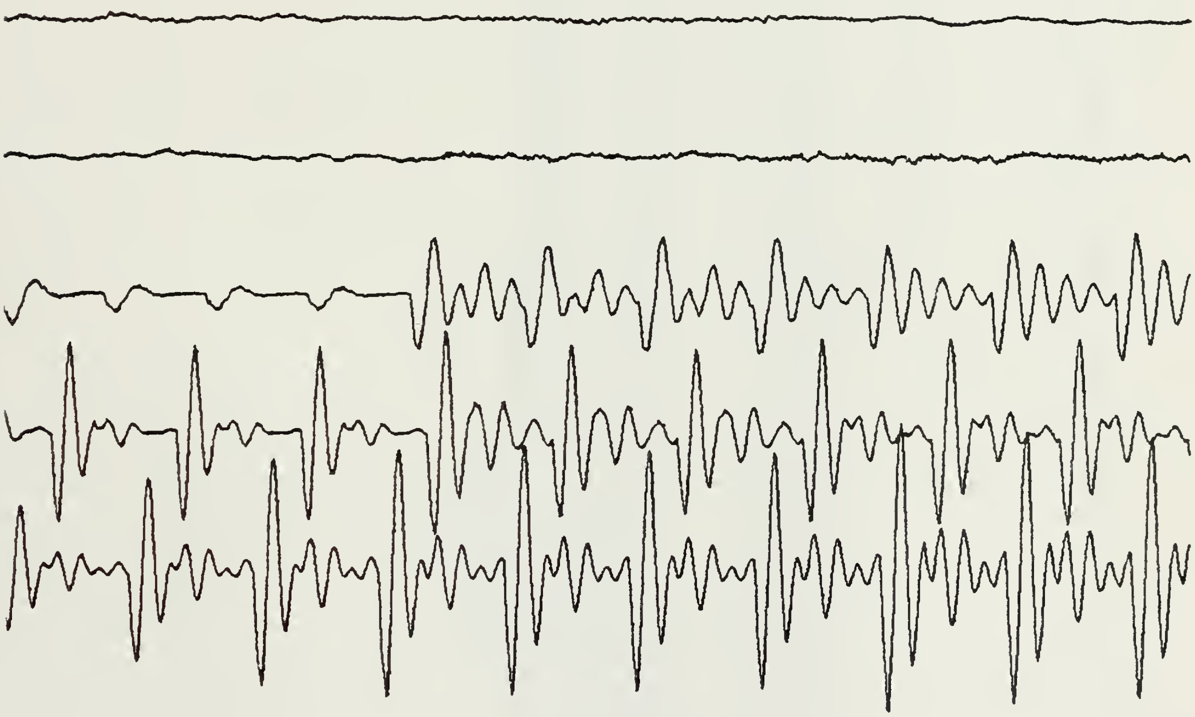


Figure B.2.8 WAVEFORM OF MODIFIED OUTPUT SPEECH

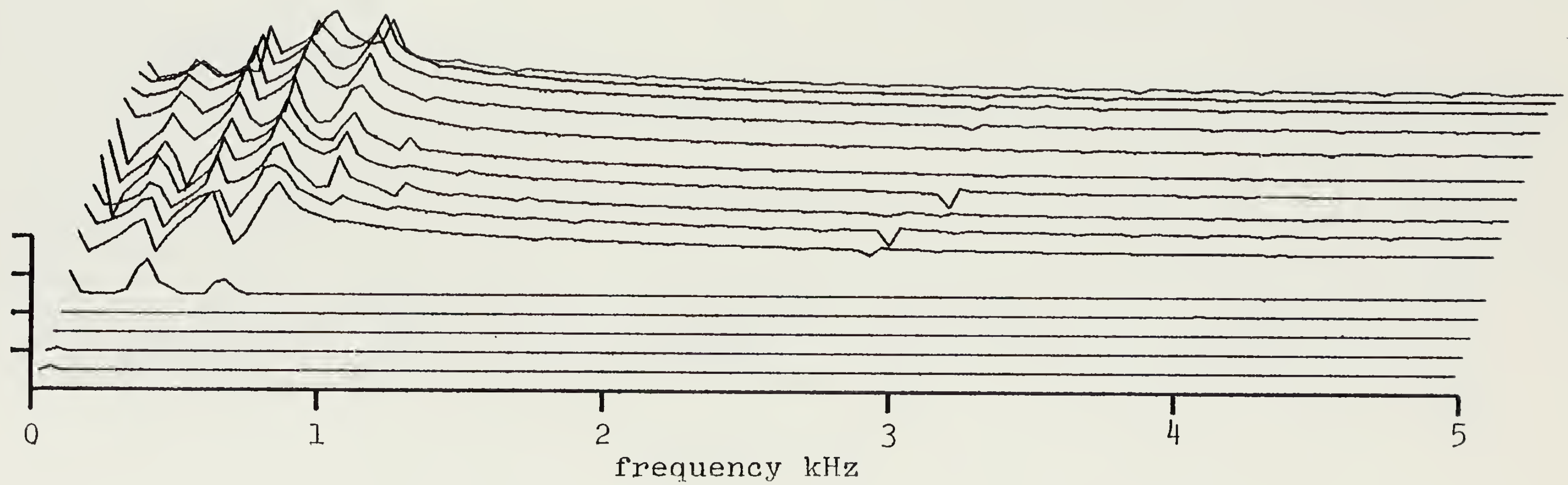


Figure B.2.9 LOGARITHMIC POWER SPECTRAL DENSITY OF UNMODIFIED OUTPUT SPEECH

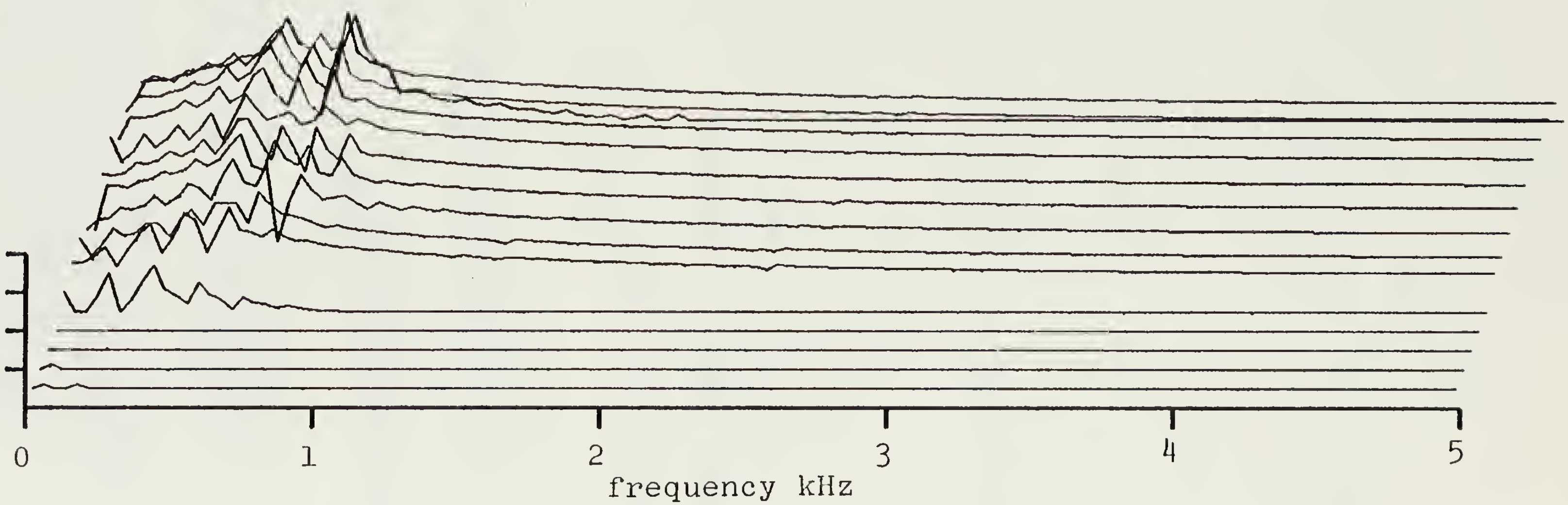


Figure B.2.10 LOGARITHMIC POWER SPECTRAL DENSITY OF MODIFIED OUTPUT SPEECH

APPENDIX C DESCRIPTION OF VOICE TAPE

The audio recording which is available from the author has four sections each of which contains three segments of speech. These three speech segments are of the following sounds:

Segment 1 - Five long vowels.

"a e i o u"

Segment 2 - Four words which are combinations of fricatives and voiced sounds.

"sat free hip done"

Segment 3 - A sentence with a varity of sounds.

"Every salt breeze comes from the sea."

Each of these segments is repeated in each segment of the tape. Each section of the tape shows the effects of a different step in the processing.

Section 1 - Unprocessed speech, the recording used for input to the processing system.

Section 2 - Speech which has been converted to digital form and then converted back to analog form with no other processing.

Section 3 - Speech which has been encoded into a set of LPC parameters and then decoded using the same parameters (i.e. no modification).

Section 4 - Speech which has been encoded into a set of LPC parameters and those parameters altered to reduce the pitch frequency by a factor of 0.56 and to reduce the formant frequencies by a factor of 0.88. The same LPC decoding process is then used to reconstruct the speech segment.

BIBLIOGRAPHY

1. Atal, B.S. and Schroeder, M.R., "Adaptive Predictive Coding of Speech Signals," Bell System Technical Journal, p. 1973-1986, October 1970.

2. Atal, B.S. and Haunaur, S.L., "Speech Analysis and Synthesis by Linear Prediction of the Speech Wave," Journal of the Acoustical Society of America, v. 50, p. 637-655, 1971.

3. Blankenship, P.E., LDVT: High Performance Minicomputer for Real-Time Speech Processing, paper presented at IEEE Electronics and Areospace Systems Convention, Washington, D. C., 29 September - 1 October 1975.

4. Makhoul, J., "Spectral Analysis of Speech by Linear Prediction," IEEE Transactions on Audio and Electroacoustics, v. AU-21, p. 140-148, June 1973.

5. Makhoul, J., "Linear Prediction: A Tutorial Review," Proceedings of the IEEE, v. 63, p. 561-580, April 1975.

6. Markel, J.D. and Gray, A.H., "A Linear Prediction Vocoder Simulation Based upon the Autocorrelation Method," IEEE Transactions on Acoustics, Speech and Signal Processing, v. ASSP-22, p. 124-134. April 1974.

7. Markel, J.D. and Gray, A.H., Linear Prediction of Speech, Springer-Verlag, 1976.

8. Massachusetts Institute of Technology Lincoln Laboratory Report 1976-37, Microprocessor Realization of a Linear Predictive Vocoder, by E. M. Hofstetter and others, 30 September 1976.

9. Rabiner, L.R. and Schafer, R.W., Digital Processing of Speech Signals, Prentice-Hall, 1978.

10. Rader, C.M., "An Improved Algorithm for High Speed Autocorrelation With Application to Spectral Estimation," IEEE Transactions on Audio and Electroacoustics, v. AU-18, p. 439-441, December 1970.

11. Schafer, R.W. and Rabiner, L.R., "Digital Representation of Speech Signals," Proceedings of the IEEE, April 1975.

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Documentation Center Cameron Station Alexandria, Virginia 22314	2
2.	Library, Code 0142 Naval Postgraduate School Monterey, California 93940	2
3.	Department Chairman, Code 52 Department of Electrical Engineering Naval Postgraduate School Monterey, California 93940	2
4.	Professor Sidney R. Parker, Code 52Px Department of Electrical Engineering Naval Postgraduate School Monterey, California 93940	5
5.	Capt. Geoffrey T. Hall, USMC 816 McBryde Drive Blacksburg, Virginia 24060	2

Thesis 179836
H14778 Hall
c.1
Computer modeling of voice signals with adjustable pitch and formant frequencies.